

(19) Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) Publication number:

0 615 201 A2

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number: 94103922.4

(51) Int. Cl.5: G06F 15/403

(22) Date of filing: 14.03.94

(30) Priority: 12.03.93 JP 79149/93

(43) Date of publication of application:  
14.09.94 Bulletin 94/37

(84) Designated Contracting States:  
DE FR GB

(71) Applicant: KABUSHIKI KAISHA TOSHIBA  
72, Horikawa-cho  
Saiwai-ku  
Kawasaki-shi Kanagawa-ken 210 (JP)

(72) Inventor: Sumita, Kazuo  
67-17, Sachigaoka,  
Asahi-ku  
Yokohama-shi, Kanagawa-ken (JP)  
Inventor: Hirakawa, Hideki  
1-18-24, Kachida-minami,  
Kouhoku-ku  
Yokohama-shi, Kanagawa-ken (JP)  
Inventor: Miike, Seiji  
6-21-19-205, Tsunashima-nishi,  
Kouhoku-ku  
Yokohama-shi, Kanagawa-ken (JP)  
Inventor: Mizutani, Yumi  
3-6-7, Yato-cho  
Tanashi-shi, Tokyo (JP)

(74) Representative: Lehn, Werner, Dipl.-Ing. et al  
Hoffmann, Eitle & Partner,  
Patentanwälte,  
Arabellastrasse 4  
D-81925 München (DE)

(54) Document detection system using detection result presentation for facilitating user's comprehension.

(57) A document detection system using a detection result presentation that can facilitate a user's quick comprehension of the relevancy of each detected document. In this system, a user's input for commanding a document detection is analyzed to extract keywords and viewpoints relevant to each keyword contained in the user's input, and a detection command is constructed from the keywords and the viewpoints extracted from the user's input. Then, the detection operation is executed to detect those documents among a plurality of stored documents which match with the constructed detection command as detected documents for a detection result. The detection result can be presented in the multi-dimensional display formed by setting the viewpoints to axes with the detection command as an origin and using distances of the detected documents with respect to the origin for each viewpoint as coordinates of the detected documents with respect to each axis representing each viewpoint.

EP 0 615 201 A2

BEST AVAILABLE COPY

BACKGROUND OF THE INVENTIONField of the Invention

5 The present invention relates to a document detection system for detecting desired documents from a large number of documents stored in a document database. It is to be noted that the term "retrieval" is often used in the literature of the field instead of the term "detection" used in the following description. The present specification adheres to the use of the term "detection" throughout.

10 Description of the Background Art

In recent years, due to the significant progress and spread of computers, the electronic manipulations of documents are becoming increasingly popular as in the electronic news and electronic mail systems and the CDROM publications of data sources such as dictionaries and encyclopedia that had only been available on papers, and it is expected that this trend of the electronic manipulations of documents will continue at an increasing pace in future.

15 In conjunction with such electronic manipulations of documents, much attentions have been attracted to a document detection system for detecting desired documents from a large number of documents efficiently, so as to enable the effective utilization of the documents stored in a database system in advance.

20 As a conventionally available document detection system, there has been a system which uses keywords in combination with logic operators such as AND, OR, NOT or proximity operators for specifying numbers of characters, sentences, and paragraphs that can exist between keywords, and detects a document by using a specified combination of keywords and operators as a detection key.

25 However, in such a conventional document detection system, the detection result has been informed by displaying either a number of detected documents or titles of the detected documents alone, so that in order for the user to check each of the detected documents to see if it is the desired document or not, it has been necessary for the user to read the entire content of each of the detected documents one by one, and this operation has been enormously time consuming.

30 Moreover, in the conventional document detection system, in displaying the titles of the detected documents, the titles are simply arranged in a prescribed order according to the user's query such as an order of descending similarities to the keywords used in the detection key. For this reason, it has been impossible for the user to comprehend the relative relationships among the detected documents and the level of similarity with respect to the detection command for each of the detected documents from the 35 displayed detection result, and consequently it has been difficult for the user to have an immediate impression for the appropriateness of the displayed detection result.

40 Furthermore, in the conventional document detection system, the detection scheme is limited to that in which each document as a whole is treated as a single entity, so that the document containing the desired content in the background section and the document containing the desired content in the conclusion section will be detected together in mixture. In other words, the detection result contains variety of 45 documents mixedly regardless of viewpoints in which the desired content appear in the documents. For example, if there is no interest in what had been done in the past, the detected document which matches with the given keywords in the background section will be of no use. Yet, in the conventional document detection system, the documents having different perspectives such as the document containing the desired content in the background section and the document containing the desired content in the conclusion section will not be distinguished, and the mixed presence of these documents in different perspectives makes it extremely difficult for the user to judge the appropriateness of the detection result.

45 In view of these problems, there has been a proposition for a scheme to reduce the burden on the user to read entire content of each detected document by displaying only a portion of each detected document. 50 However, in such a scheme, it is often impossible to make a proper judgement as to whether it is the desired document or not unless the relationship of the displayed portion and the remaining portion becomes apparent. For example, when the background section containing the desired content is displayed for one document while the conclusion section containing the desired content for the other document, as these 55 documents cannot be comprehended in a unified viewpoint, it is difficult for the user to make a proper judgement as to which one of these document is the necessary one. As a result, in order to fully comprehend the perspectives of the displayed portions in these documents, the user would be forced to read the entire contents of these documents after all, so that it cannot contribute to the reduction of the burden on the user at all.

Also, there has been a proposition for a scheme to reduce the burden on the user to read the entire content of each detected document by providing a man-made document summary for each stored document in advance in correspondence to each stored document itself and displaying the document summary at a time of displaying the detection result. However, in such a scheme, an enormous amount of 5 human efforts is required for preparing the document summary for each document at a time of producing the database itself, which is not practically justifiable unless the database system has a remarkably high utilization rate. Moreover, there are many already existing database systems in which the document summary for each document is not provided, and an enormous amount of human efforts is similarly required for preparing the document summary for each document in such an already existing database 10 system. In addition, the man-made document summary is produced in the very general viewpoint alone, so that there is no guarantee that each document is summarized from a viewpoint suitable for the required detection. As a result, the document summary displayed as the detection result can be quite out of point from the viewpoint of the user with the specific document detection objective, and in such a case, it is possible for the user to overlook the actually necessary document at a time of judging whether each 15 detected document is the desired document or not.

#### SUMMARY OF THE INVENTION

It is therefore an object of the present invention to provide a document detection system using a 20 detection result presentation that can facilitate a user's quick comprehension of the relevancy of each detected document, so as to enable the user to conduct an entire operation using the document detection system smoothly.

According to one aspect of the present invention there is provided a document detection system, comprising: document memory means for storing a plurality of documents; input means for entering user's 25 input for commanding a document detection in the documents stored in the document memory means; input analysis means for analyzing the user's input entered by the input means to extract keywords and viewpoints relevant to each keyword contained in the user's input, and constructing a detection command from the keywords and the viewpoints extracted from the user's input; and detection means for detecting those documents stored in the document memory means which match with the detection command 30 constructed by the input analysis means as detected documents of a detection result.

According to another aspect of the present invention there is provided a method of document detection, comprising the steps of: analyzing a user's input for commanding a document detection to extract keywords and viewpoints relevant to each keyword contained in the user's input; constructing a detection command from the keywords and the viewpoints extracted from the user's input at the analyzing step; and detecting 35 those documents among a plurality of stored documents which match with the detection command constructed at the constructing step as detected documents of a detection result.

Other features and advantages of the present invention will become apparent from the following description taken in conjunction with the accompanying drawings.

#### 40 BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic block diagram for an overall hardware configuration of a first embodiment of a document detection system according to the present invention.

Fig. 2 is a block diagram for a detailed functional configuration of the document detection system of 45 Fig. 1.

Fig. 3 is a block diagram of a detailed functional configuration of an input analysis unit in the document detection system of Fig. 2.

Fig. 4 is a flow chart for the operation of an input analysis control unit in the input analysis unit of Fig. 3.

Fig. 5 is a flow chart for the operation of a detection command generation unit in the input analysis unit 50 of Fig. 3.

Fig. 6 is a table of exemplary viewpoint extraction rules used in a viewpoint extraction unit in the input analysis unit of Fig. 3.

Fig. 7 is an illustration of a format and examples for a detection command generated by the detection command generation unit in the input analysis unit of Fig. 3.

Fig. 8 is a block diagram of a detailed functional configuration of a detection unit in the document detection system of Fig. 2.

Fig. 9 is a flow chart for the detection operation of a detection control unit in the detection unit of Fig. 8.

Fig. 10 is a diagrammatic illustration of a keyword index stored in a document data memory unit in the document detection system of Fig. 2.

Fig. 11 is an illustration of a format and examples for a detection result obtained by the detection unit of Fig. 8.

5 Fig. 12 is a flow chart for the re-detection operation of a detection control unit in the detection unit of Fig. 8.

Fig. 13 is a block diagram of a detailed functional configuration of a record management unit in the document detection system of Fig. 2.

10 Fig. 14 is a flow chart for the operation of a record management control unit in the record management unit of Fig. 13.

Fig. 15 is an illustration of a format and an example for a detection node stored in a detection record memory unit in the document detection system of Fig. 2.

Fig. 16 is a flow chart for the operation of a detection record display unit in the document detection system of Fig. 2.

15 Figs. 17A and 17B are illustrations of exemplary detection record displays presented by the detection record display unit in the document detection system of Fig. 2.

Fig. 18 is a flow chart for the operation of a detection result display unit in the document detection system of Fig. 2.

20 Fig. 19 is illustrations of exemplary detection result displays presented by the detection result display unit in the document detection system of Fig. 2.

Fig. 20 is an illustration of a format and examples for an accept/reject data used in the document detection system of Fig. 2.

Fig. 21 is a flow chart for the operation of a browsing unit in the document detection system of Fig. 2.

25 Fig. 22 is a diagrammatic illustration of a data format for the data content of the document stored in the document data memory unit in the document detection system of Fig. 2.

Fig. 23 is illustrations of exemplary browsing unit displays presented by the browsing unit in the document detection system of Fig. 2.

Figs. 24A and 24B are a flow chart for the modified operation of the detection result display unit in a second embodiment of a document detection system according to the present invention.

30 Fig. 25 is schematic illustrations of exemplary detection result displays presented by the detection result display unit according to the flow chart of Figs. 24A and 24B.

Figs. 26A and 26B are a flow chart for the modified operation of the detection result display unit in a third embodiment of a document detection system according to the present invention.

35 Figs. 27A, 27B, and 27C are schematic illustrations of exemplary detection result displays presented by the detection result display unit according to the flow chart of Figs. 26A and 26B.

Fig. 28 is an illustration of an exemplary detection record display presented by the detection record display unit in a fourth embodiment of a document detection system according to the present invention.

Fig. 29 is an illustration of a window for detection command sentence input used in a fifth embodiment of a document detection system according to the present invention.

40 Fig. 30 is a flow chart for the modified operation of the detection command generation unit in a sixth embodiment of a document detection system according to the present invention.

Fig. 31 is an illustration of a viewpoint extraction result presented in the sixth embodiment of a document detection system according to the present invention.

45 Fig. 32 is a flow chart for the modified operation of the record management control unit in the sixth embodiment of a document detection system according to the present invention.

Fig. 33 is an illustration of a format and an example for a detection node stored in the detection record memory unit in the sixth embodiment of a document detection system according to the present invention.

Fig. 34 is a flow chart for the modified operation of the detection record display unit in the sixth embodiment of a document detection system according to the present invention.

50 Figs. 35A and 35B are illustrations of exemplary detection record displays presented by the detection record display unit in the sixth embodiment of a document detection system according to the present invention.

Fig. 36 is an illustration of the detection node data presented by the detection record display unit in the sixth embodiment of a document detection system according to the present invention.

55 Fig. 37 is a flow chart for the modified operation of the detection result display unit in the sixth embodiment of a document detection system according to the present invention.

Fig. 38 is a flow chart for the clusterization at the detection result display unit in the flow chart of Fig. 37.

- Fig. 39 is a flow chart for the clusterization at the detection result display unit in the flow chart of Fig. 38.
- Fig. 40 is an illustration of a format and an example for a cluster document data presented by the detection result display unit in the sixth embodiment of a detection system according to the present invention.
- Fig. 41 is an illustration of an exemplary detection result display in a cluster display mode presented by the detection result display unit in the sixth embodiment of a detection system according to the present invention.
- Fig. 42 is illustrations of exemplary detection result displays in a document display mode presented by the detection result display unit in the sixth embodiment of a detection system according to the present invention.
- Fig. 43 is an illustration of an exemplary detection result display in a document display mode presented by the detection result display unit in the sixth embodiment of a detection system according to the present invention.
- Figs. 44A and 44B are illustrations of exemplary browsing unit displays presented by the browsing unit in the sixth embodiment of a detection system according to the present invention.
- Fig. 45 is a diagrammatic illustration of a data format for the data content of the document stored in the document data memory unit in the sixth embodiment of a detection system according to the present invention.
- Fig. 46 is an illustration of an exemplary detection result display in a cluster display mode presented by the detection result display unit in the seventh embodiment of a detection system according to the present invention.
- Fig. 47 is an illustration of an exemplary detection result display in a document display mode presented by the detection result display unit in the seventh embodiment of a detection system according to the present invention.
- Fig. 48 is an illustration of an exemplary detection result display presented by the detection result display unit in the eighth embodiment of a detection system according to the present invention.
- Fig. 49 is an illustration of an exemplary table indicating similarities among detected documents presented by the detection result display unit in the eighth embodiment of a detection system according to the present invention.
- Fig. 50 is an illustration of an exemplary table indicating viewpoints utilized in the document data memory unit in the ninth embodiment of a detection system according to the present invention.
- Fig. 51 is a table of exemplary numerical value expression extraction rules used in the tenth embodiment of a detection system according to the present invention.
- Fig. 52 is an illustration of an exemplary detection result display presented by the detection result display unit in the tenth embodiment of a detection system according to the present invention.
- Figs. 53A and 53B are illustrations of exemplary detection result displays presented by the detection result display unit in the eleventh embodiment of a detection system according to the present invention.
- Fig. 54 is a table of exemplary viewpoint extraction rules used in the twelfth embodiment of a detection system according to the present invention.
- Figs. 55A, 55B, 55C, and 55D are illustrations of windows for registering a viewpoint extraction rule in the twelfth embodiment of a detection system according to the present invention.
- Figs. 56A and 56B are illustrations of exemplary detection record displays presented by the detection record display unit in the thirteenth embodiment of a document detection system according to the present invention.
- Fig. 57 is an illustration of another exemplary detection record display presented by the detection record display unit in the thirteenth embodiment of a document detection system according to the present invention.
- Fig. 58 is an illustration of an exemplary browsing unit display presented by the browsing unit in the fourteenth embodiment of a document detection system according to the present invention.
- Fig. 59 is an illustration of exemplary windows for browsing unit displays presented by the browsing unit in the fourteenth embodiment of a document detection system according to the present invention.
- Fig. 60 is a diagrammatic illustration of a data format for the individual data used in the fourteenth embodiment of a detection system according to the present invention.
- Fig. 61 is an illustration of an exemplary detection result display in a document display mode presented by the detection result display unit in the fifteenth embodiment of a detection system according to the present invention.

Fig. 62 is an illustration of another exemplary detection result display in a document display mode presented by the detection result display unit in the fifteenth embodiment of a detection system according to the present invention.

Fig. 63 is an illustration of exemplary windows for browsing unit displays presented by the browsing unit in the sixteenth embodiment of a document detection system according to the present invention.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring now to Fig. 1, the first embodiment of a document detection system according to the present invention will be described in detail.

In this first embodiment, the document detection system has an overall hardware configuration as shown in Fig. 1 in which a central processing means 101 is connected with a memory means 102, a display means 104 through a display controller 103, and an input means 106 through an input controller 105.

The central processing means 101 is formed by a processor for carrying out various processing operations. The memory means 102 is formed by a memory medium such as a semiconductor memory, a magnetic disk memory, an optical disk memory, etc. for storing programs and data used by the central processing means 101. The display means 104 is formed by a display device such as a liquid crystal display and a plasma display for displaying the text content of the document and the detection result under the control of the display controller 103. The input means 106 is formed from input devices such as a keyboard and a mouse for entering the detection commands from the user under the control of the input controller 106.

In further detail, the document detection system of this first embodiment has a detailed functional configuration as shown in Fig. 2, which comprises an input unit 201, an input analysis unit 202, a detection unit 203, a record management unit 204, a detection record memory unit 205, a document data memory unit 206, a detection result display unit 207, a detection record display unit 208, and a browsing unit 209, which are mutually connected by data lines 210-227 indicated by thick solid lines and control lines 228-238 indicated by thin solid lines in Fig. 2, as will be described in detail below.

The input unit 201 enters input sentences from the user given in a natural language or keywords for commanding the document detection operation. The input sentences are subsequently converted into the detection command at the input analysis unit 202 and the converted detection command is returned to this input unit 201, via the data lines 210 and 211. The input unit 201 receives a detection node ID from the record management unit 204 via the data line 213, and outputs the detection node ID and the detection command in pair to the record management unit 204 via the data line 212 and to the detection unit 203 via the data line 214.

The input analysis unit 202 receives the user's input from the input unit 201 via the data line 210, converts the user's input into the detection command, and outputs the converted detection command to the input unit 201 via the data line 211.

The detection unit 203 receives the pair of the detection command and the detection node ID from the input unit 201 via the data line 214, looks up document data stored in the document data memory unit 206 via the data line 218 to detect the related document set, and outputs a detection result concerning the related document set and the detection node ID in pair to the record management unit 204 via the data line 215, while outputting the detection result to the detection result display unit 208 via the data line 216.

The detection result display unit 207 receives the detection result via the data line 216, displays the document set of the detection result as a multi-dimensional display, and outputs an document ID of a document selected in the multi-dimensional display to the browsing unit 209 via the data line 227.

The browsing unit 209 looks up the document data stored in the document data memory unit 206 via the data line 219 according to the document ID received from the detection result display unit 207 via the data line 227, to display the content of the individual document.

The record management unit 204 stores the pair of the detection command and the detection node ID received via the data line 212 and the pair of the detection result and the detection node ID received via the data line 215 into the detection record memory unit 205 via the data line 223, while looking up the detection record via the data line 222, and outputting the detection record data to the detection record display unit 207 via the data line 225.

The detection record display unit 208 displays the detection record data received via the data line 225 in a tree structure, and outputs the detection node ID of a node specified in the displayed tree structure to the record management unit 204 via the data line 224.

The detection record memory unit 205 stores the detection record received from the record management unit 204 via the data line 223, while the document data memory unit 206 stores the document data.

The data line 226 is for transmitting an accept/reject data specified for each individual document by the user on the browsing unit 209 to the detection result display unit 207, the data line 220 is for transmitting the detection node ID generated at the record management unit 204 to the detection result display unit 207, the data lines 221 and 217 are for transmitting the immediately previous detection result, the document 5 accept/reject data, and the detection node ID at a time of re-detection specified at the detection result display unit 207 to the detection unit 203 and the record management unit 204.

The input analysis unit 202 has a detailed functional configuration as shown in Fig. 3, which comprises an input analysis control unit 301, a detection command generation unit 302, a viewpoint extraction unit 303, a morphological analysis unit 304, and a syntactic analysis unit 305, which are mutually connected by data 10 lines 308-315 indicated by thick solid lines and control lines 317-320 indicated by thin solid lines in Fig. 3, as will be described in detail below.

The input analysis control unit 301 is activated via the control line 228 from the input unit 201. This input analysis control unit 301 activates the morphological analysis unit 304 via the control line 319 and supplies the individual input sentence to the morphological analysis unit 304 via the data line 312, and then receives 15 the morphological analysis result from the morphological analysis unit 304 via the data line 313. In addition, this input analysis control unit 301 activates the syntactic analysis unit 305 via the control line 320 and supplies the morphological analysis result of the individual input sentence to the syntactic analysis unit 305 via the data line 314, and then receives the syntactic analysis result from the syntactic analysis unit 305 via the data line 315. Furthermore, this input analysis control unit 301 activates the detection command 20 generation unit 302 via the control line 317 and supplies the syntactic analysis result of the individual input sentence to the detection command generation unit 302 via the data line 308, and then receives the detection command from the detection command generation unit 302 via the data line 309.

The detection command generation unit 302 activates the viewpoint extraction unit 303 via the control line 318 for each individual input sentence and supplies the syntactic analysis result via the data line 310, 25 and then receives the viewpoint data extracted by the viewpoint extraction unit 303 via the data line 311. In addition, this detection command generation unit 302 extracts content words from the individual input sentence, constructs the detection command from the extracted content words, and outputs the constructed detection command to the input analysis control unit 301 via the data line 309.

The input analysis control unit 301 operates according to the flow chart of Fig. 4 as follows.

First, there is a possibility for a plurality of natural language sentences to be entered from the input unit 30 201. For this reason, with respect to each individual input sentence entered, the morphological analysis and the syntactic analysis are carried out at the steps 401 and 402 by using the morphological analysis unit 304 and the syntactic analysis unit 305. Here, the morphological analysis and the syntactic analysis are already well known from the field of machine translation, and the details of these analyses are not essential to the 35 present invention so that their explanation will be omitted.

Then, according to the syntactic analysis result obtained for each individual input sentence entered, the detection command generation is carried out at the step 403 by using the detection command generation unit 302.

The detection command generation unit 302 operates according to the flow chart of Fig. 5 as follows.

40 Namely, according to the syntactic analysis result for individual input sentence, the viewpoint data is extracted at the step 501, and the content words constituting the individual input sentence are extracted at the step 502.

Then, after these viewpoint extraction and the content word extraction are carried out for all the input 45 sentences, the extracted content words are merged for each viewpoint at the step 503, so as to construct the detection command.

Here, the extraction of the viewpoint data is carried out according to the viewpoint extraction rules such as those shown in a table of Fig. 6 which only shows a limited number of examples. In Fig. 6, each viewpoint extraction rule is given in a format of:

(matching portion) → viewpoint  
50 where the matching portion indicates a syntactic pattern to be matched, such that when the syntactic pattern of the matching portion on the left hand side matches with that in the individual input sentence, the viewpoint on the right hand side is determined as the viewpoint of the individual input sentence.

In a case the viewpoint data cannot be extracted according to the viewpoint extraction rules of Fig. 6, the detection command generation unit 302 constructs the detection command by adopting the default 55 viewpoint set in advance. This default viewpoint may be set up from the input unit 201 before the execution of the detection command by the user.

For example, in a case of the viewpoint extraction rule at the first line in Fig. 6, a matching is made for a partial element of a sentence containing the syntactic pattern of "with an object of", while in a case of the

viewpoint extraction rule at the second line in Fig. 6, a matching is made for a partial element of a sentence containing the syntactic pattern of "understand [perfect tense]", i.e., "understood", "have understood", etc.

It is to be noted however that the format of this matching portion depends on the syntactic analysis result obtained by the syntactic analysis and any suitable format other than that used in the examples of Fig. 6 may be used.

In the content word extraction, the noun parts are taken out from the syntactic analysis result, and then the unnecessary words are removed according to an unnecessary word dictionary (not shown) in the already known manner of the content word extraction from the syntactic analysis result.

The detection command has an exemplary format as shown in Fig. 7, where the detection command is defined as a pair of the keyword and the viewpoint list. The examples shown in Fig. 7 are four detection commands constructed from the input sentence of "A design tool based on examples developed with an object of a computer design.", in which the nouns "computer" and "design" are subordinate to "with an object of" while the nouns "example", "design", and "tool" are subordinate to "developed".

The detection unit 203 has a detailed functional configuration as shown in Fig. 8, which comprises a detection control unit 801 for controlling the detection unit 203 as a whole, an index search unit 802 for detecting related documents by searching through a keyword index stored in the document data memory unit 206, an adaptation unit 803 for receiving the accept/reject data for the documents from the detection result display unit 207 and carrying out an adaptation operation in a case of the redetection, and a distance calculation unit 804 for calculating the similarities of the detected documents with respect to the detection commands and converting the calculated similarities into distances.

The detection control unit 801 carries out two types of operations including a detection operation when the detection commands are received and a re-detection operation when the accept/reject data for the documents are received from the detection result display unit 207.

The detection operation by the detection control unit 801 is carried out according to the flow chart of Fig. 9 as follows.

This detection operation is carried out first when the detection control unit 801 is activated via the control line 230.

First, the index search unit 802 is activated via the control line 812 according to the detection commands received via the data line 214. At this point, the detection commands are transmitted to the index search unit 802 via the data line 813.

In response, the index search unit 802 searches through the keyword index, and obtains the document IDs and the viewpoint data containing the keywords of the detection commands as candidate documents at the step 901. The document IDs and the viewpoint data obtained at the step 901 are then supplied to the detection control unit 801 via the data line 807.

Then, for all the candidate documents obtained at the step 901, the distance calculation unit 804 is activated via the control line 811, and the calculations of distances with respect to the detection commands are carried out at the step 902. At the distance calculation unit 804, the keywords of each candidate document and the detection commands are received from the detection control unit 801 via the data line 808, and the distance calculation is carried out according to an  $M \times N$  matrix representation of each detection command  $Q$  defined by the following equation (1), a  $M \times N$  matrix representation of each document  $D_i$  defined by the following equation (2), and the distance  $Dist(Q, D_i)$  defined by the following equation (3). The obtained distance is subsequently supplied to the detection control unit 801 via the data line 809.

$$Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} & \cdots & q_{1M} \\ q_{21} & q_{22} & q_{23} & \cdots & q_{2M} \\ q_{31} & q_{32} & q_{33} & \cdots & q_{3M} \\ | & | & | & & | \\ q_{N1} & q_{N2} & q_{N3} & \cdots & q_{NM} \end{bmatrix} \quad (1)$$

$$5 \quad D_i = \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1M} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2M} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3M} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{NM} \end{pmatrix} \quad (2)$$

10       $Dist(Q, D_i) = |Q - D_i|/M \quad (3)$

where the ij-th elements  $q_{ij}$  and  $d_{ij}$  at the i-th row and j-th column in the matrices of the equations (1) and (2) are for the i-th keyword and the j-th viewpoint of the detection command and the document, respectively, 15 which express weights for the keywords which are determined by analysing the detection command and the document.

For example, the ij-th element  $q_{ij}$  has a value 1 when the i-th keyword is used in the sentence belonging to the j-th viewpoint, or a value 0 otherwise. Here, when the keywords are developed into synonyms by using the synonym dictionary (not shown), it is also possible to give a value less than 1 and 20 greater than 0 to the inferior words, superior words, or synonymous words. It is also possible to give a value less than 1 and greater than 0 according to the position of the keyword within the document structure such as the title, the chapter header, the main text, the footnote, etc.

These matrices Q and Di can be regarded as expressing feature vectors of the detection command and the document with respect to the viewpoints.

25      In the above equation (3), a symbol  $|A|$  for an arbitrary matrix A represents the meaning defined by the following equation (4).

$$30 \quad |A| = \begin{pmatrix} \sum b_{1m} \\ \sum b_{2m} \\ \sum b_{3m} \\ \vdots \\ \sum b_{Nm} \end{pmatrix} \quad (4)$$

35      where, for each element  $a_{ij}$  of the matrix A, if  $a_{ij} < 0$  then  $b_{ij} = -a_{ij}$ , and otherwise  $b_{ij} = a_{ij}$ .

40      It is to be noted that the distance  $Dist(Q, D_i)$  of the above equation (3) is the city block distance, but any other generally known distance measures may be used instead, if desired.

By the above equation (3), the distance between the detection command and the individual document is going to be obtained for each viewpoint.

45      Fig. 10 shows an exemplary content of the keyword index stored in the document data memory unit 206, which is suitable for Japanese.

Namely, this keyword index of Fig. 10 has the TRIE structure in which each kanji character involved in the keywords is assigned to a unique address, and each keyword formed by a plurality of kanji characters is specified by the link data registered after each kanji character, so as to reduce the required memory capacity and simplify the necessary detection procedure.

50      For example, the Japanese keyword "kikai" (meaning "machine") formed by two kanji characters, the first character registered at the address 00305 in the head character storage region has a link data "00935" specifying the second character of the keyword as that registered at the address 00305 in the subsequent character storage region. In addition, this second character at the address 00305 also has the link data "00623" specifying the third character and the third character at the address 00623 has the link data "00914" specifying the fourth character, for another keyword "kikai-honyaku" (meaning "machine translation") formed by four kanji characters which contains the above described keyword "kikai" as a first part.

Furthermore, the second character at the address 00305 also has a file data "file 4" indicating that the keyword "kikai" is contained in the document data having the document ID "file 4", accompanied by a

viewpoint data "conclusion" indicating that this keyword "kikai" is used in the sentence concerning the viewpoint of "conclusion" in this document data "file 4".

Similarly, the fourth character at the address 00914 has two pairs of file data and viewpoint data set "file 25, object" and "file 21, conclusion", indicating that the keyword "kikai-honyaku" is contained in the 5 document data "file 25" in the sentence concerning the viewpoint of "object", and in the document data "file 21" in the sentence concerning the viewpoint of "conclusion".

On the other hand, the first character at the address 000A0 in the head character storage region is common to two keywords "sanpo" (meaning "algorithm") and "sanjutsu" (meaning "arithmetic"), so that it has two link data "00A15" and "00A16" specifying the respective second characters for these two 10 keywords.

In Fig. 10, an isolated "0" functions as a separator for separating the character, link data, and file data and viewpoint data pair. Also, the first characters of the keywords are registered in the continuous head character storage region in a sorted order such as that of JIS (Japanese Industrial Standard) codes.

Fig. 11 shows an exemplary format for detection result along with the actual examples of the detection 15 results.

In this case, each document ID is associated with a plurality of distances, where the ij-th distance  $\langle\text{DISTANCE}_{ij}\rangle$  is for the i-th document  $\langle\text{DOCUMENT}_i\rangle$  and the j-th viewpoint.

The detection results obtained by the detection unit 203 are then supplied to the record management unit 204 via the data line 215, as well as to the detection result display unit 207 via the data line 216.

20 On the other hand, the re-detection operation by the detection control unit 801 is carried out according to the flow chart of Fig. 12 as follows.

In this case, this re-detection operation is activated by the detection result display unit 207 via the control line 232. The detection control unit 801 then activates the adaptation unit 803 via the control line 810, using the detection result containing the accept/reject data for the documents as the input via the data 25 line 217, so as to re-construct the detection commands at the step 1201. Here, the adaptation unit 803 calculates each re-constructed detection command  $Q'$  from each previous detection command  $Q$  according to the following equation (5).

$$30 \quad Q' = w_0 \cdot Q + \sum_{k=1}^K w_k \cdot Q_k \quad (5)$$

where  $Q_k$  indicates the matrix representation of the document which is judged as proper for the detection 35 result, i.e., accepted in the accept/reject data, and  $w_k$  indicates a weight for taking a weighted average.

After this re-calculation of the detection commands at the step 1201, the steps 1202 and 1203 similar to the steps 901 and 902 in Fig. 9 for the detection operation described above are carried out for the re-constructed detection commands  $Q'$ .

The record management unit 204 has a detailed functional configuration as shown in Fig. 13, which 40 comprises a record management control unit 1301 and a detection generation unit 1302.

The record management control unit 1301 operates according to the flow chart of Fig. 14 as follows.

Here, in accordance with an input message, the operation selectively proceeds to an appropriate one of the following operations of the steps 1402, 1043, and 1405 as described below from the step 1401.

Namely, in a case the the generation of the detection node ID is commanded from the input unit 201 45 via the control line 231, or from the detection result display unit 207 via the control line 233, the generation of the detection node using the detection node generation unit 1302 is carried out at the step 1402. The generated detection node is then stored into the detection record memory unit 205 via the data line 223.

In this case, each detection node is given in a format as shown in Fig. 15 along with an example, which 50 comprises a set of four elements including the detection node ID, the parent node ID, the detection commands, and the detection results. Here, the detection node ID stores an identifier (ID) for identifying this detection node while the parent ID stores the ID of the detection node which stores the detection commands and the detection results for the immediately previous detection operation. The record management unit 204 is going to manage the detection record according to this detection node ID.

In this generation of the detection node, a new detection node is generated by setting the current 55 detection node as the parent node. Then, the detection node ID of the newly generated detection node is outputted to the input unit 201 via the data line 213 or to the detection result display unit 207 via the data line 220.

On the other hand, in a case a pair of the detection command and the detection node ID is entered from the input unit 201 via the data line 212 and a pair of the detection result and the detection node ID is entered from the detection unit 203 via the data line 215, the storing of the detection command and the detection result is carried out at the step 1403.

- 5 Then, the detection record display unit 208 is commanded via the control line 236 to transfer the detection nodes stored in the detection record memory unit 205 to the detection record display unit 208 via the data line 225, so as to update the detection record display at the step 1404.

Also, in a case the change of the current detection node is commanded from the detection record display unit 208 via the control line 237, the detection node specified by using the mouse or other input device on the detection record display unit 208 is entered via the data line 224, and the current detection node ID is changed accordingly at the step 1405. Then, the detection record display is updated at the step 1406, while the detection result display unit 207 is commanded to accordingly update the detection result display via the control line 238 at the step 1407.

The detection record display unit 208 operates according to the flow chart of Fig. 16 as follows.

- 15 Here, in accordance with an input message, the operation selectively proceeds to an appropriate one of the following operations of the steps 1602, 1604, 1605, 1606, and 1607 as described below from the step 1601.

First, the generation of a display window for this detection record display unit 208 is commanded from the input unit 201 via the control line 235. In this case, the generation of a new window for the detection record display is carried out at the step 1605. Also, when the deletion of the window for the detection record display is commanded from the input unit 201 via the control line 235, the deletion of the window for the detection record display is carried out at the step 1606.

- 20 In a case a new record display is commanded from the record management unit 204 via the control line 236, the detection record to be newly displayed is read out via the data line 225 at the step 1602, and the detection record display is updated by the read out detection record on the window for the detection record display at the step 1603. Here, an exemplary detection record display made by this detection record display unit 208 is shown in Fig. 17A.

As shown in Fig. 17A, the detection record display is given in a form of the tree structure, in which the detection nodes are represented by black dots, where two nodes linked by a straight line have the parent and child relationship, and an encircled black dot indicates the current detection node. Here, in addition, the current detection node is displayed in a display mode for clearly distinguishing it from the other nodes such as blinking display mode, inverted display mode, distinct color display mode, etc.

At this detection record display unit 208, the user inputs for commanding the change of the current detection node and the generation of a new window for the detection result display can be received.

- 35 In a case of changing the current detection node, the user can enter the input for specifying the desired detection node among those displayed on the window for the detection record display by using the mouse or other input device. When the change of the current detection node is commanded, the detection record display unit 208 commands the record management unit 204 to set the specified detection node as a new current detection node, in response to which the record management control unit 1301 carries out the steps 40 1405 to 1407 of Fig. 14 described above. By changing the current detection node in the exemplary detection record display of Fig. 17A, the detection record display can be changed as shown in Fig. 17B, for example.

On the other hand, in a case the user input for generating a new window for the detection result display is received, the detection record display unit 208 commands the detection result display unit 207 to generate and display a new window for the detection result display via the control line 239.

The detection result display unit 207 operates according to the flow chart of Fig. 18 as follows.

- 45 Here, in accordance with an input message, the operation selectively proceeds to an appropriate one of the following operations of the steps 1802, 1803, 1804, 1806, 1807, and 1810 as described below from the step 1801.

50 Namely, this detection result display unit 207 carries out one of the display of the detection results entered from the detection unit 203 via the data line 216, the display of the detection results in the detection record data entered from the record management unit 204 via the data line 220, the input of the accept/reject data for the documents from the browsing unit 209 via the data line 226, the generation of the window for the detection result display commanded from the detection record display unit 208 via the control line 239, and the generation of the window for the detection result display commanded from the input unit 201 via the control line 229.

In addition, this detection result display unit 207 also carries out one of the change of the displayed viewpoint, the activation of the browsing unit 209, and the re-activation of the detection unit 203 according

to the accept/reject data for the documents, upon receiving the user's input.

In a case the generation of the window for the detection result display is commanded from the input unit 201 or the detection record display unit 208, the window for the detection result display is generated at the step 1802. Here, the command from the detection record display unit 208 comes from the step 1607 of Fig. 16.

Also, when the detection results are entered from the detection unit 203 via the data line 216, or the detection results in the detection record data are entered from the record management unit 204 via the data line 220, the entered detection results are displayed at the step 1803.

Here, this detection result display is given in a form of a multi-dimensional display with each viewpoint stored in the detection results as an axis, the detection commands as an origin, and the distances between the detection commands and the candidate documents for each viewpoint as coordinates of data points representing the candidate documents.

In this first embodiment, this multi-dimensional display is given in an exemplary form shown in a part (a) of Fig. 19, which is showing a three-dimensional case as an example. Here, each axis is labelled by the viewpoint represented by each axis, such as "object" and "conclusion" for example.

In this example of Fig. 19, one of the viewpoint axis is labelled "others" because in this example a plurality of viewpoint axes are mapped onto this one axis labelled "others" as a number of viewpoints extracted from the detection commands exceeds a predetermined simultaneously displayable viewpoint number, so that the multi-dimensional space displayed by Fig. 19 is a contracted space obtained from the actual higher dimensional space. Here, the coordinates along this axis labelled "others" can be calculated by taking a weighted average of the coordinates along the axes to be contracted according to the following equation (6), for example.

$$25 \quad r = \sum_i v_i \cdot r_i \quad (6)$$

where  $r$  is a coordinate along the axis labelled "others",  $r_i$  is a coordinate along an  $i$ -th axis to be contracted, and  $v_i$  is a weight for the  $i$ -th axis to be contracted.

30 It is also possible to modify the above equation (6) as in the following equation (7).

$$35 \quad r = \sqrt{\sum_i v_i \cdot (r_i)^2} \quad (7)$$

In general, the formula for calculating the coordinates along this axis labelled "others" can be expressed as the following equation (8).

$$40 \quad r = \text{func2}(\sum_i v_i \cdot \text{func1}(r_i)) \quad (8)$$

45 where  $\text{func1}$  and  $\text{func2}$  are suitable functions.

In stead of displaying the label "others" as shown in Fig. 19, it is also possible to use the label explicitly indicating the viewpoints represented by the contracted axes, such as "background + method" for example.

It is also to be noted that the three-dimensional display of the example shown in Fig. 19 can be easily reduced to the two-dimensional one or the one dimensional one by further contracting the axes.

In the part (a) of Fig. 19, the data points for the documents are indicated by black dots, accompanied by the document titles indicated in the solid line enclosures.

In a case the user entered the accept/reject data for the documents currently browsed at the browsing unit 209, the accept/reject data are received via the data line 226 at the step 1804. Here, each accept/reject data is received in the format of:

$\langle\text{document ID}\rangle;\langle\text{accept/reject}\rangle$

where  $\langle\text{accept/reject}\rangle$  is indicated by a value "0" indicating the reject of the document identified by the accompanying document ID as improper for the detected document or a value "1" indicating the accept of

this document as proper for the detected document.

Then, at the detection result display unit 207, the display of the document corresponding to the entered document ID is updated accordingly as shown a part (c) of Fig. 19 at the step 1805. In this part (c) of Fig. 19, the data points for those documents which are rejected by the accept/reject data are indicated by blank circles, while the data points for those documents which are accepted by the accept/reject data remain to be indicated by black dots, and the document titles accompanying the blank circles and the black dots are displayed in different colors, as indicated by the dashed line enclosures and the solid line enclosures in the part (c) of Fig. 19.

Here, the accept/reject data received from the browsing unit 209 are stored in the data format shown in Fig. 20, until the new detection results are received.

The detection result display unit 207 can also activate the browsing unit 209 in response to the user's input using the mouse or the other input device on the window for the detection result display.

Namely, when the data point representing the individual document is specified on the window for the detection result display by the user, the detection result display unit 207 activates the browsing unit 209 for the specified document via the control line 234 at the step 1806. In this case, the detection result display is changed as shown in the part (b) of Fig. 19 in which the document title accompanying the data point for the currently browsed document is displaying in the blinking display mode or in the distinct color display mode, as indicated by the shading within the solid line enclosure in the part (b) of Fig. 19.

In a case a displayed button "re-detect" is selected by the user on the window for the detection result display, the detection unit 203 is reactivated at the step 1807 according to the current detection results including the accept/reject data for currently displayed documents supplied via the data line 217.

In a case a displayed button "viewpoint" is selected by the user on the window for the detection result display, the viewpoint is changed to a new viewpoint at the step 1808. Here, in changing the viewpoint, which viewpoint is to be allocated to which axis is changed in accordance with the user's input. Then, the detection results are displayed from the new viewpoint at the step 1809.

In a case a displayed button "generate" is selected by the user on the window for the detection result display, the step 1802 described above is carried out again to generate additional window for the detection result display newly. Then, on the generated additional window, the same detection results as those displayed on the original window are also displayed. Here, by selecting different viewpoints for the original and additional windows, the detection results can be displayed from two different viewpoints. Here, however, the detection results themselves and the accept/reject data are shared among these original and additional windows, so that the display of the accept/reject data for the documents are going to be identical in the original and additional windows.

In a case a displayed button "delete" is selected by the user on the window for the detection result display, the window for the detection result display is deleted at the step 1810.

The browsing unit 209 operates according to the flow chart of Fig. 21 as follows.

Here, in accordance with an input message, the operation selectively proceeds to an appropriate one of the following operations of the steps 2102, 2104, 2108, 2115, and 2116 as described below from the step 2101.

This browsing unit 209 is activated from the detection result display unit 207 via the control line 234 as described above, and displays the data content of the document specified by the user's input.

Namely, in a case this browsing unit 209 is activated by the step 1806 of Fig. 18, the window for the browsing unit display is generated at the step 2102. At this point, the document ID of the specified document is also entered from the detection result display unit 207 via the data line 227. Then, the document data of the specified document is read out from the document data memory unit 206 via the data line 219 and the summary of the specified document is displayed at the step 2103 on the window generated at the step 2102. Here, instead of the summary, the original document may be displayed if desired.

Here, the data content of the document is stored in the document data memory unit 206 in an exemplary data format shown in Fig. 22, in which pointers to original document, summary, keyword list, viewpoint list, and document structure are collectively registered.

In this case, the browsing unit display can be given in an exemplary form as shown in Fig. 23, in which a part (a) shows an initial window display showing the summary of the specified document in a case of the step 2103 described above. This initial window display also provides displayed buttons for "viewpoint" and "keyword" and displayed buttons "OK" and "NG" for entering the accept/reject data.

In a case a displayed button "viewpoint" is selected by the user on the window for the browsing unit display, the viewpoint list enlisting all the viewpoints stored in the document data for the specified document is presented in a menu format shown in a part (c) of Fig. 23 at the step 2104, and the selection

of the desired viewpoint by the user is awaited at the step 2105. Then, unless it is the abort at the step 2106, the text content of the document concerning the specified viewpoint is displayed in a form shown in a part (e) of Fig. 23 at the step 2107.

In a case a displayed button "keyword" is selected by the user on the window for the browsing unit display, the keyword list enlisting all the keywords stored in the document data for the specified document is presented in a menu format shown in a part (b) of Fig. 23 at the step 2108, and the selection of the desired keyword by the user is awaited at the step 2109. Then, unless it is the abort at the step 2110, the viewpoint list enlisting all the viewpoints concerning the specified keyword is presented in a menu format shown in a part (d) of Fig. 23 at the step 2111, and the selection of the desired viewpoint by the user is awaited at the step 2112. Then, unless it is the abort at the step 2113, the text content of the document concerning the specified viewpoint is displayed in a form shown in a part (e) of Fig. 23 at the step 2114.

In a case a displayed button "OK" or "NG" is selected by the user on the window for the browsing unit display to specify the accept/reject data, the entered accept/reject data is transmitted to the detection result display unit 207 at the step 2115.

In a case the deletion of the window for the browsing unit display is commanded by the user, the window for the browsing unit display is deleted at the step 2116.

As described in detail, according to this first embodiment, the distances between the detection commands and detected documents are obtained for a plurality of viewpoints, and the detection results are presented by displaying the obtained distances on the multi-dimensional space formed by using the viewpoints as axes, such that the user can readily comprehend how each detected document is close to the detection commands in what viewpoint according to the distribution of the data points on the multi-dimensional display.

Thus, according to this first embodiment, it is possible to provide a document detection system using a detection result presentation that can facilitate a user's quick comprehension of the relevancy of each detected document, so as to enable the user to conduct an entire operation using the document detection system smoothly.

It is also possible in this first embodiment to keep a record of the detection result display before the re-detection operation by storing the detection results and the corresponding accept/reject data as the detection record data, such that the detection result display before the re-detection operation can be reproduced even after the re-detection operation.

It is to be noted that the keyword index of Fig. 10 used in the first embodiment may be provided in plurality in correspondence to the plurality of viewpoints, such that the viewpoint data registered in each TRIE structure entry of the keyword index of Fig. 10 can be omitted.

It is also to be noted that the keyword index of Fig. 10 given in the TRIE structure may be replaced by the keyword index using the other known high speed referencing scheme such as that using the hash function.

Now, the second embodiment of a document detection system according to the present invention will be described in detail.

This second embodiment is a modification of the first embodiment described above concerning the operation of the detection result display unit 207. The features other than the operation of the detection result display unit 207 are substantially equivalent as those of the first embodiment described above.

Namely, in this second embodiment, the detection result display unit 207 operates according to the flow chart of Figs. 24A and 24B as follows.

First, whether an input message is a mouse input from the user or not is judged at the step 2402. If not, in accordance with the input message, the operation selectively proceeds to an appropriate one of the following operations of the steps 2402, 2403, 2404, and 2405 as described below from the step 2401. Here, the step 2403, 2404, 2405, and 2406 are equivalent to the steps 1802, 1803, 1804, and 1805 in the flow chart of Fig. 18 for the first embodiment, so that their descriptions will be omitted.

In a case of the mouse input, if the mouse input is a "delete" command, a "generate" command, a "re-detect" command, or a "viewpoint" command is sequentially judged at steps 2407, 2409, 2411, and 2413. When it is one of these commands, the operation proceeds to the steps 2408, 2410, 2412, or 2414 and 2415, respectively, which are equivalent to the steps 1810, 1802, 1807, or 1808 and 1809, respectively, in the flow chart of Fig. 18 for the first embodiment, so that their descriptions will be omitted. In addition, whether the mouse input is a clicking of a point indicating the document or not is judged at the step 2416. If so, the operation proceeds to the step 2417 which is equivalent to the step 1806 in the flow chart of Fig. 18 for the first embodiment, so that its description will be omitted.

Next, whether the mouse input is a clicking of an end of an axis or not is judged at the step 2418. If so, the axis specified by the clicking is rotated along the movement of the mouse entered by the user while

appropriately enlarging or contracting the axis at the step 2419. For example, starting from the display as shown in a part (a) of Fig. 25, one axis can be rotated as shown in a part (b) of Fig. 25, or an origin can be shifted by rotating the axes together as shown in a part (c) of Fig. 25. Then, using the axes in changed orientations, the detection result is re-displayed at the step 2420.

- 5      Otherwise, whether the mouse input is a "enlarge" command or not is judged at the step 2421. If so, a range specification for a portion to be enlarged is entered by the user as shown in a part (d) of Fig. 25 at the step 2422. Then, the enlarged display is re-displayed by applying a normalization in the specified range. Namely, the enlarged display is obtained by re-scaling the distances according to the following equation (9).

10

$$15 \quad \text{Dist}'(Q, D_i) = \begin{bmatrix} b_1' \\ b_2' \\ b_3' \\ | \\ b_N' \end{bmatrix} \quad (9)$$

20

where  $b_j' = a_j/R_j$ ,  $a_j$  is an element of  $\text{Dist}(Q, D_i)$ , and  $R_j$  is a range of each axis obtained by a range specification.

- Otherwise, whether the mouse input is a "contract" command or not is judged at the step 2424. If so, a range specification for a portion to be contracted is entered by the user at the step 2425. Then, the contracted display is re-displayed by setting the specified range as an entire size. Namely, the contracted display is obtained by re-scaling the distances according to the following equation (10).

30

$$35 \quad \text{Dist}''(Q, D_i) = \begin{bmatrix} b_1'' \\ b_2'' \\ b_3'' \\ | \\ b_N'' \end{bmatrix} \quad (10)$$

35

where  $b_j'' = a_j \times R_j$ ,  $a_j$  is an element of  $\text{Dist}(Q, D_i)$ , and  $R_j$  is a range of each axis obtained by a range specification.

- 40      Otherwise, whether the mouse input is a "accept/reject" command or not is judged at the step 2427. If so, a range specification for a portion within which the accept/reject data of the documents are to be inverted is entered by the user at the step 2428. Then, the display in which the accept/reject data for each document in the specified range is inverted is re-displayed at the step 2429.

It is to be noted that the above steps 2421 to 2427 may be modified such that the range specification can be made first, before the command content is specified.

- 45      It is also to be noted that in the detection result display by the detection result display unit 207, the perspective with respect to the origin in the depth direction can be expressed by the overwriting the data point and the title of the document located on a closer side over the data point and the title of the document located on a farther side. Moreover, the other computer graphic techniques for expressing the three-dimensional perspective such as a color gradation or a shading may be employed in addition, if desired.

50      It is also to be noted that, after the last step of each operation sequence in this flow chart of Fig. 24, the next input event such as the mouse input is awaited, and the operation starts from the beginning in response to the next input event.

- Now, the third embodiment of a document detection system according to the present invention will be described in detail.

This third embodiment is a further modification of the second embodiment described above concerning the operation of the detection result display unit 207. The features other than the operation of the detection result display unit 207 are substantially equivalent as those of the first embodiment described above.

Namely, in this third embodiment, the detection result display unit 207 additionally operates according to the flow chart of Fig. 26A as follows.

Here, before the step 2618, the detection result display unit 207 can carry out the operations similar to those of the steps 2401 to 2417 of Fig. 24 for the second embodiment.

5 Then, at the step 2618, whether the mouse input is a "rotate" command or not is judged. If so, the input event is awaited at the step 2619, and whether the next input event is a clicking of a vicinity of an axis or not is judged at the step 2620.

If so, each data point representing each document of the detection result is rotated by a prescribed angle  $\theta$  around the axis specified by the mouse input, and the new coordinates of each data point after the 10 rotation are obtained at the step 2621. Then, the detection result is displayed by using the new coordinates obtained at the step 2621 for each data point at the step 2622.

Here, the display of the detection result at the step 2622 can be carried out according to the flow chart of Fig. 26B, in which the coordinates of each data point representing each document of the detection result are projected onto a prescribed two-dimensional plane at the step 2623, and then the detection result 15 display is obtained according to the projected coordinates of each data point on the two-dimensional plane. This detection result display operation of Fig. 26B can also be employed at the other steps requiring the detection result display such as those of the steps 2404, 2415, and 2420 in the flow chart of Fig. 24 for the second embodiment.

In the detection result display operation of Fig. 26B, the projection onto the two-dimensional plane is 20 necessary obviously because of the need to present the multi-dimensionaoal detection result display on a two dimensional display screen.

To this end, the general method for projecting a point in a certain dimensional coordinate system into a lower dimensional coordinate system is mathematically well known as a projection onto a partial space. The projection onto the two-dimensional plane at the step 2623 corresponds to a special case of this general 25 method, so that this well known general method can be utilized at the step 2623.

As for the rotation at the step 2621, it is also well known that the coordinate vector "a" of each point can be transformed into a rotated coordinate vector "a'" by a rotation around a z-axis by an angle  $\theta$  according to the following equation (11).

$$30 \quad a' = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

35

Thus, in a case of the steps 2621 and 2622 in Fig. 26A, after the coordinates of all the data points are transformed according to this equation (11), the projection onto the two dimensional plane can be carried out to obtain the detection result display after the rotation.

40 Here, the general rotation around an arbitrary axis can be expressed as a linear combination of the rotations around three orthogonal coordinate axes, so that the rotation around any given axis can be obtained by sequentially carrying out the rotation around each of the orthogonal coordinate axes.

As an example, when the detection result display shown in Fig. 27A is rotated by selecting the "rotate" button and clicking a vicinity of the vertical axis in Fig. 27A, the detection result display can be changed as 45 shown in Fig. 27B in which the entire detection result display is rotated around the vertical axis for a prescribed angle.

In addition, in obtaining the detection result result by the operation of Fig. 26B by projecting the multi-dimensional detection result onto the two dimensional plane, the three-dimensional perspective can be enhanced by providing supplementary line segments as shown in Fig. 27C, where the supplementary line 50 segments for each data point represent line segments extended from each data point along x, y, and z axes until they intersect with the y-z plane, z-x plane, and x-y plane, respectively, defined by the coordinate axes indicated by the solid lines.

It is to be noted that the mouse input used in this third embodiment for the purpose of specifying the rotation operation may be replaced by any other input device capable of the same function, such as a data globe utilized in the three dimensional object display manipulation in the computer graphics for example.

55 Next, the fourth embodiment of a document detection system according to the present invention will be described in detail.

In this fourth embodiment, the detection record display in a form of a tree structure used in the first embodiment described above is modified to further incorporate the contracted display of the corresponding detection result in a vicinity of each detection node as shown in Fig. 28.

With this modified detection record display of Fig. 28, it becomes easier for the user to comprehend the past detection performances.

Here, the contracted display of the detection result can be obtained by the manner similar to that used in the second embodiment in response to the mouse input for the "contract" command.

Next, the fifth embodiment of a document detection system according to the present invention will be described in detail.

10 In this fifth embodiment, instead of extracting the viewpoint data from the syntactic analysis result of the input sentences and using the default viewpoint setting in a case of the failure to extract any viewpoint data from the syntactic analysis result of the input sentences as in the first embodiment described above, it is made possible to carry out the setting of the viewpoint interactively, in a menu format, by providing a window for detection command sentence input equipped with a button for a viewpoint selection, as shown in Fig. 29. In this case, when the user selects the viewpoint button, the viewpoint menu enlisting all the available viewpoints appears as indicated in Fig. 29, such that the user can select any desired viewpoint from this viewpoint menu. Here, the viewpoint set up immediately before the detection command sentence input is finished by entering the last return is going to be employed for the construction of the detection command.

20 In the example shown in Fig. 29, a detection command sentence of "sophistication of knowledge" is entered from a viewpoint of "object", while another detection command sentence of "development of expert system" is entered from a viewpoint of "conclusion", as indicated at a right edge field in the window. This window may also have a detection button along the viewpoint button, as indicated in Fig. 29.

25 Next, the sixth embodiment of a document detection system according to the present invention will be described in detail.

In this sixth embodiment, the first embodiment described above are expanded in several aspects as follows. Here, only those features which are different from the corresponding features in the first embodiment will be described in detail.

30 In this sixth embodiment, the detection command generation unit 302 in the input analysis unit 202 operates according to the flow chart of Fig. 30, in which the steps 3001, 3002, and 3003 which are identical to the steps 501, 502, and 503 of Fig. 5 for the first embodiment are followed by a step 3004 in which a window for a viewpoint extraction result display is generated, and a step 3005 in which the viewpoint extraction result is displayed on the window generated at the step 3004.

35 In this case, the viewpoint extraction itself is carried out similarly to the first embodiment, but the result of the viewpoint extraction is displayed in a form shown in Fig. 31, in which the keywords associated with each viewpoint are tabulated.

Also, in this case, the distance calculation unit 804 of the detection unit 203 receives the viewpoint data along with the keywords and the detection commands from the detection control unit 801 via the data line 808.

40 On the other hand, the record management control unit 1301 of the record management unit 204 operates according to the flow chart of Fig. 32 as follows.

Namely, in this flow chart of Fig. 32, the steps 3201 to 3207 are identical to the steps 1401 to 1407 in Fig. 14 for the first embodiment. In addition, there is provided a step 3208 between the steps 3206 and 3207 in which the detection node data are updated and displayed, and a step 3209 that can be proceeded from the step 3201 in which the accept/reject data for the documents are stored.

50 In this case, each detection node is given in a format as shown in Fig. 33 along with an example, where, in addition to the detection node ID, the parent node ID, the detection commands, and the detection results included in Fig. 15 for the first embodiment, Fig. 33 further includes a detection date and time, a detected input sentence, a number of detected documents, a number of OK documents, a list of the OK documents, a number of NG documents, and a list of the NG documents. Here, entries concerning the detection reflects the detection operation, while entries concerning the OK documents and the NG documents indicates the accept/reject data for the documents entered by the user at the browsing unit 209.

The detection record display unit 208 displays the detection record display in a form of the tree structure shown in Fig. 35A, which differs from that of Fig. 17A in that each detection node is represented by a black dot is accompanied by the detection date on left side and the number of detected documents on right side.

More specifically, the detection record display unit 208 operates according to the flow chart of Fig. 34 as follows.

Namely, in this flow chart of Fig. 34, the steps 3401 to 3407 are identical to the steps 1601 to 1607 in Fig. 16 for the first embodiment. In addition, there is provided a step 3408 in which the detection node data are read out, followed by a step 3409 in which the read out detection node data are updated and displayed. Moreover, there are a step 3410 in which a window for the detection node data display is generated, a step 5 3411 in which a window for the detection node data display is generated, and a step 3412 in which a generation and a display of the window for the browsing unit display are commanded in a case the user specified the OK document name or the NG document name.

Here, the detection node data is displayed on the window for the detection node data display at the step 3411 in an exemplary form as shown in Fig. 36 in correspondence to the detection node of Fig. 33.

10 In this sixth embodiment, the detection result display unit 207 operates according to the flow chart of Fig. 37 as follows.

Namely, in this flow chart of Fig. 37, the steps 3701 to 3710 are identical to the steps 1801 to 1810 in Fig. 18 for the first embodiment.

In addition, in this sixth embodiment, there is a prescribed threshold for clusterizing to decide whether 15 or not to clusterize the detection result display in which the documents within a prescribed distance range are grouped together to form a cluster to be displayed collectively. In a case a total number of detected documents does not exceeds this prescribed threshold, the multi-dimensional display of the detection result is made at the step 3703 just as in the first embodiment.

On the other hand, when the total number of detected documents exceeds the prescribed threshold, the 20 detection result is clusterized at the step 3712, and then the clusterized detection result is displayed at the step 3713.

Here, the clusterization at the step 3712 is carried out according to the flow chart of Fig. 38 as follows. Namely, first the detected documents of the detection result are clusterized such that a number of detected documents per each cluster becomes less than the prescribed threshold for clusterizing at the step 3801. 25 Then, whether a number of clusters resulting from the step 3801 is less than a prescribed minimum number of clusters or not is determined at the step 3802. When the number of clusters resulting from the step 3801 is less than the prescribed minimum number of clusters at the step 3802, a cluster having a largest number of detected documents is further clusterized at the step 3803 and the operation returns to the step 3802.

In further detail, the clusterization at the step 3801 is carried out according to the flow chart of Fig. 39 30 as follows. Namely, first the coordinate space of the detection result display is divided into eight sub-spaces at planes parallel to two coordinate axes and passing through a center of another one coordinate axes at the step 3901. Then, whether a number of detected documents in any sub-space is greater than a prescribed threshold for clusterizing is determined at the step 3902. When there is a sub-space for which the number of detected documents is greater than the prescribed threshold at the step 3902, this sub-space containing 35 an excessive number of detected documents is further divided into eight sub-spaces at the step 3903, and the operation returns to the step 3902.

In this clusterization operation of Fig. 39, the cluster document data shown in Fig. 40 for indicating the correspondences between the clusters and the detected documents are generated, to keep track of the clusterization operation.

40 In this case, the clusterized detection result is displayed on the window for detection result display in an exemplary cluster display mode shown in Fig. 41, where each cluster is displayed according to a distance with respect to the detection commands at an origin in a three-dimensional space formed by axes representing the viewpoints. Here, each cluster is represented as a small sphere located at a coordinate position calculated from the coordinate positions of the detected documents contained in each cluster, such 45 as an average coordinate position of the coordinate positions of the detected documents contained in each cluster for example. It is also possible to display a title of an arbitrary detected document contained in each cluster along the small sphere representing each cluster as a representative title.

In addition, the number of detected documents contained in each cluster is expressed by changing a size or a concentration level of each small sphere in accordance with the number of detected documents 50 contained in each cluster. It is also possible to indicate the number of detected documents contained in each cluster by numerals to be displayed inside or in a vicinity of the small sphere representing each cluster.

It is also to be noted that in the detection result display in the cluster display mode of Fig. 41 by the detection result display unit 207, the perspective with respect to the origin in the depth direction can be 55 expressed by the overwriting the small sphere for the cluster located on a closer side over the small sphere for the cluster located on a farther side. Moreover, the other computer graphic techniques for expressing the three-dimensional perspective such as a color gradation or a shading may be employed in addition, if desired. It is also possible to assign a serial number for the clusters in an increasing order of the distance

with respect to the origin, and indicate this serial number for each cluster by numerals to be displayed inside or in a vicinity of the small sphere representing each cluster.

In addition, the detection result display in the cluster display mode of Fig. 41 also contains an indication of the total number of the detected documents, and displayed buttons for "viewpoint", "develop", "cluster", "generate", and "delete".

When the displayed button "cluster" is selected, a window for a cluster data display is generated at the step 3714 of Fig. 37, and the cluster data indicating a number of clusterized documents, a number of detected documents per cluster, and a number of clusters are displayed in this window at the step 3715, along with a displayed button "re-display" as shown in Fig. 41. Here, the user can change the settings of these numbers freely and command the re-display by selecting "re-display" button to obtain the re-display of the detection result display in the cluster display mode according to the changed settings. Also, by selecting the displayed button "cluster" again in this state, the window for the cluster data display can be deleted by the step 3716 of Fig. 37.

On the other hand, when the displayed button "develop" is selected while specifying at least one cluster by the mouse or other input device on the window for the detection result display, a new window for the detection result display is generated at the step 3702, and the multi-dimensional display for the detected documents within the specified cluster alone is displayed at the step 3703, in a form of an enlarged display in which the coordinate range of the specified cluster is set as an entire display coordinate space.

Here, in a case the user failed to specify the cluster properly, it is possible to display the group of clusters located within a predetermined distance from the specified point together in the enlarged display. Alternatively, it is also possible to display the cluster located closest to the specified point in the enlarged display. It is further possible to display the detected documents contained in the specified cluster sequentially in a prescribed order on the browsing unit 209. It is also possible to change the representative title of the detected document displayed in a vicinity of the small sphere representing each cluster sequentially in a prescribed order among the detected documents contained in each cluster.

Moreover, independently from the user's specification of the cluster, the detected documents of a prescribed number of clusters in an order of the distance with respect to the origin can be displayed on the separate windows for the detection result display, or the detected documents of a prescribed number of clusters in an order of a number of the detected documents contained in each cluster can be displayed on the separate windows for the detection result display. In these case, it is also possible to indicate the correspondences between the small spheres representing clusters and the windows for the detection result display which are displaying the detected documents of these clusters.

In this case, the window for the detection result display can display the detected documents of each cluster in response to the selection of the displayed button "develop" in an exemplary document display mode shown in a part (a) of Fig. 42, which is similar to the entire detection result display of Fig. 19 in the first embodiment, except that each axis representing each viewpoint has an indication for a range of that viewpoint for the displayed cluster. In addition, the this detection result display in the document display mode of Fig. 42 also contains indications of the total number of the detected documents, and a number of the detected documents in the displayed cluster, and displayed buttons for "viewpoint", "re-detect", "generate", "delete", and "arrange".

It is also to be noted that in this window for the detection result display in the document display mode of Fig. 42 provided by the detection result display unit 207, the perspective with respect to the origin in the depth direction can be expressed by the overwriting the data point and the title of the document located on a closer side over the data point and the title of the document located on a farther side. Moreover, the other computer graphic techniques for expressing the three-dimensional perspective such as a color gradation or a shading may be employed in addition, if desired. It is also possible to assign a serial number for the detected documents in an increasing order of the distance with respect to the origin, and indicate this serial number for each detected document by numerals to be displayed in a vicinity of the data point representing each detected document.

In a case of using no clusterization as a total number of detected documents does not exceeds the prescribed threshold, the multi-dimensional display of the detection result at the step 3703 can be made similarly, except that the number of the detected documents in the displayed cluster is not included in such a case.

In the detection result displays of Figs. 41 and 42, the scale on each viewpoint axis is given as a real scale at constant intervals, but it is also possible to utilizes a logarithmic scale if desired. It is also possible to allow the user to select the desired scale. In an example of Fig. 41, the scale runs from 0 to 100, while in an example of Fig. 42, the scale runs from 0 to 50 corresponding to the range covered by the coordinate

space for the displayed cluster. Here, it is also possible to set the scale range in Fig. 42 in correspondence to the distance of the detected document or the cluster which is farthest from the origin along each axis within the displayed cluster or cluster group.

When one of the displayed buttons "viewpoint", "generate", or "delete" is selected on the detected result display in the cluster display mode of Fig. 41, the steps 3708 and 3709, 3702, or 3710, respectively, are carried out similarly to the first embodiment.

Similarly, when one of the displayed buttons "viewpoint", "re-detect", "generate", or "delete" is selected on the detected result display in the document display mode of Fig. 42, the steps 3708 and 3709, 3707, 3702, or 3710, respectively, are carried out similarly to the first embodiment.

Also, the detection result display in the document display mode shown in a part (a) of Fig. 42 can be changed to those shown in a part(b) or a part (c) of Fig. 42 in response to the activation of the browsing unit 209 or the accept/reject data entered from the browsing unit 209, just as in the cases of parts (b) and (c) of Fig. 19 in the first embodiment. In this case, the accept/reject data for the documents are memorized until the detection operation is terminated, at which point the accept/reject data for the documents are transmitted to the detection record management unit 204 and stored for the current detection node by the step 3717 of Fig. 37.

On the other hand, when the displayed button "arrange" is selected on the detected result display in the document display mode of Fig. 42, the browsing unit 209 is activated for each of the detected documents specified up to that point, and the windows for the browsing unit display are arranged in an order specified by the user at the step 3711 of Fig. 37.

In this case, in response to the selection of the displayed button "arrange", a menu in an exemplary form shown in Fig. 43 appears, from which the manner of arranging the windows for the browsing unit display can be selected. In this example of Fig. 43, the user can choose any one of "distance from origin" arrangement in an increasing order of the distances of the specified documents with respect to the origin, "document production date" arrangement in an increasingly getting older order of the dates for the productions of the specified documents, "author name (alphabetical)" arrangement in an alphabetical order of the authors of the specified documents, and "access date & time" arrangement in an increasingly getting older order of the latest accesses to the specified documents for the purpose of the display.

In response to the selection by the user on the menu of Fig. 43, an original state of the windows for the browsing unit display as shown in Fig. 44A can be changed into an arranged state as shown in Fig. 44B. In this display of the windows for the browsing unit display of Figs. 44A and 44B, it is also possible to change the color of the concentration level of the frame or the background of each window according to a particular order, such as an increasing order of the distances of the specified documents with respect to the origin, an increasingly getting older order of the dates for the productions of the specified documents, an alphabetical order of the authors of the specified documents, or an increasingly getting older order of the latest accesses to the specified documents for the purpose of the display.

For the purpose of this arrangement of the windows for the browsing unit display, the data content of the document stored in the document data memory unit 206 in this sixth embodiment further registers a pointer to time data containing the production date and the access data & time information as shown in Fig. 45, in addition to the format of Fig. 22 for the first embodiment.

Next, the seventh embodiment of a document detection system according to the present invention will be described in detail.

In this seventh embodiment, the detection result display unit 207 of the sixth embodiment described above is modified such that the accept/reject data for the documents can be entered on the detection result display in the cluster display mode of Fig. 41 or the detection result display in the document display mode of Fig. 42 provided by the detection result display unit 207, as follows.

Namely, in this seventh embodiment, the detection result display in the cluster display mode is given in a form shown in Fig. 46 which contains a displayed button "document" instead of the displayed button "develop" in Fig. 41.

When this displayed button "document" is selected by the user after specifying at least one cluster, a menu containing items "develop", "OK", and "NG" appears as shown in Fig. 46. When the item "develop" is selected on this menu, the operation for displaying the detection result in the document display mode is carried out by the step 3703 of Fig. 37 just as in a case of selecting the displayed button "develop" in the sixth embodiment.

On the other hand, when the item "OK" is selected in this menu, the accept/reject data for all the detected documents contained in the specified cluster are set to 1 indicating the acceptance, whereas when the item "NG" is selected in this menu, the accept/reject data for all the detected documents contained in the specified cluster are set to 0 indicating the rejection.

On the window for the detection result display in the cluster display mode shown in Fig. 46, the small spheres representing those clusters for which the item "develop" or "OK" has been selected are shown by black spheres, while the small spheres representing those clusters for which the item "NG" has been selected are shown as white spheres, such that the state of the accept/reject data for each cluster can be indicated by the color of the small spheres representing the clusters. It is also possible to use the different concentration levels of the same color for indicating the state of the accept/reject data for each cluster. The accept/reject data entered on this window for the detection result display in the cluster display mode are subsequently transmitted to the record management unit 204 by the step 3717 of Fig. 37.

Similarly, in this seventh embodiment, the detection result display in the document display mode is given in a form shown in Fig. 47 which contains a displayed button "select" in addition to the displayed buttons contained in Fig. 42.

When this displayed button "select" is selected by the user after specifying at least one document, a menu containing items "develop", "OK", and "NG" appears as shown in Fig. 47. When the item "develop" is selected on this menu, the operation for displaying the data content of the specified document on the window for the browsing unit display is carried out by the step 3706 of Fig. 37 just as in a case of specifying the document on the window for the detection result display in the document display mode in the sixth embodiment.

On the other hand, when the item "OK" is selected in this menu, the accept/reject data for the specified detected document is set to 1 indicating the acceptance, whereas when the item "NG" is selected in this menu, the accept/reject data for the specified detected document is set to 0 indicating the rejection.

On the window for the detection result display in the document display mode shown in Fig. 47, the data points and titles representing those detected documents for which the item "develop" or "OK" has been selected are shown in black, while the data points and titles representing those detected documents for which the item "NG" has been selected are shown in white, such that the state of the accept/reject data for each detected document can be indicated by the color of the data points and titles representing the detected documents. It is also possible to use the different concentration levels of the same color for indicating the state of the accept/reject data for each detected document. The accept/reject data entered on this window for the detection result display in the document display mode are subsequently transmitted to the record management unit 204 by the step 3717 of Fig. 37.

The remaining features of this seventh embodiment are substantially identical to those of the sixth embodiment described above.

It is to be noted that the detection result displays of Figs. 46 and 47 may be given in one or two-dimensional display instead of being in the three-dimensional display as shown. It is also possible to modify the manner of specifying the clusters and the detected documents such that the user can specify the range on the display in order to specify all the clusters or the detected documents contained within the specified range collectively.

Next, the eighth embodiment of a document detection system according to the present invention will be described in detail.

In this eighth embodiment, the manner of presenting the detection result in the sixth embodiment described above is modified as follows.

Namely, in this eighth embodiment, instead of presenting the detection result display in the cluster display mode given in the multi-dimensional display as in the sixth embodiment, the number of detected documents for each cluster are indicated in a form of a two-dimensional graph as shown in Fig. 48 according to the coordinate position of each cluster, and the similarities among the detected documents of each cluster in various viewpoints are indicated in a form a table shown in Fig. 49.

In Fig. 48, the numerals shown within circles representing clusters on this two-dimensional graph indicate the number of detected documents contained in the respective clusters, whereas in Fig. 49, the numerical values enlisted below each viewpoint heading indicate the similarity levels of the enlisted detected documents in each viewpoint.

Next, the ninth embodiment of a document detection system according to the present invention will be described in detail.

In this ninth embodiment, the first embodiment described above is modified to present the distribution of the documents stored in the document data memory unit 206 according to the viewpoints, by checking all the indices within the document data memory unit 206, independently from the detection operation based on the selected viewpoints extracted from the input sentences.

Namely, in this ninth embodiment, when a plurality of document data classes such as a plurality of databases are contained in the document data memory unit 206, a number of documents stored in each of these document data classes for each of the viewpoints utilized in the document data memory unit 206 is

presented in a form of a table shown in Fig. 50, in which the total number of the documents as well as a number of the documents in each viewpoint including those already displayed as the detected documents are enlisted for each document data class identified by its name.

Also, whenever the document stored in the document data memory unit 206 is updated, it is possible to indicate the change of the distribution of the stored documents by updating the numbers enlisted in this table of Fig. 50.

Next, the tenth embodiment of a document detection system according to the present invention will be described in detail.

In this tenth embodiment, the detection operation and the multi-dimensional display of the detection result based on the selected viewpoints extracted from the input sentences in the first embodiment described above is modified such that the detection operation and the multi-dimensional display of the detection result can be carried out according to the numerical values such as date, price, and processing speed, which are extracted from the input sentences according to the prescribed numerical value expression extraction rules.

Namely, in this tenth embodiment, each numerical value expression extraction rule is given in a format of:

(matching portion) → numerical value expression type

as in the typical examples shown in a table of Fig. 51, in which the variables X, Y, Z represent the numerical values to be extracted.

In this case, the detection result display can be given in a form of a two-dimensional graph as shown in Fig. 52 for an exemplary case of using the numerical values of the price and the sales date of a certain product such as an IC memory for example. In this Fig. 52, the circles represent the documents such as the newspaper articles concerning the price and the sales date of each product.

Next, the eleventh embodiment of a document detection system according to the present invention will be described in detail.

In this eleventh embodiment, the sixth embodiment described above is modified such that the user can specify the range for each axis instead of specifying a small sphere or a data point or title representing the cluster or the document in the detection result display, such that the enlarged display of the specified range can be obtained.

Namely, on the original detection result display as shown in Fig. 53A which is given in the two-dimensional display in this example, the user can specify the ranges on two axes as indicated by dashed lines, and in response, the enlarged display of the data within this specified ranges as shown in Fig. 53B can be obtained.

Here, either the original detection result display of Fig. 53A is replaced by the enlarged display of Fig. 53B, or both of the original detection result display of Fig. 53A and the enlarged display of Fig. 53B can be presented together side by side.

Next, the twelfth embodiment of a document detection system according to the present invention will be described in detail.

In this twelfth embodiment, the viewpoint extraction rules used in the first embodiment described above is modified to incorporate different weight factors for different viewpoints.

Namely, in this twelfth embodiment, each viewpoint extraction rule is given in a format of:

(matching portion) → viewpoint, weight

as in the typical examples shown in a table of Fig. 54.

In this case, the matrix representations Q and Di of the detection command and the document are given by the following equations (12) to (15).

$$Q = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_N \end{bmatrix} \quad (12)$$

where

$$5 \quad X_j = \begin{pmatrix} q_{11} & q_{12} & q_{13} & \cdots & q_{1M} \\ q_{21} & q_{22} & q_{23} & \cdots & q_{2M} \\ q_{31} & q_{32} & q_{33} & \cdots & q_{3M} \\ | & | & | & & | \\ q_{K1} & q_{K2} & q_{K3} & \cdots & q_{KM} \end{pmatrix} \quad (13)$$

$$10 \quad D_i = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{pmatrix} \quad (14)$$

20 where

$$25 \quad Y_j = \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1M} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2M} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3M} \\ | & | & | & & | \\ d_{K1} & d_{K2} & d_{K3} & \cdots & d_{KM} \end{pmatrix} \quad (15)$$

30 and where the j-th elements  $X_j$  and  $Y_j$  at the j-th row of  $Q$  and  $D_i$  in the respective equations (12) and (14) are for the j-th viewpoint, and the ij-th elements  $q_{ij}$  and  $d_{ij}$  at the i-th row and j-th column of  $X_j$  and  $Y_j$  in the respective equations (13) and (15) are for the i-th keyword and the j-th weight of the detection command and the document, respectively.

35 Then, the formula for calculating the distance  $\text{Dist}(Q, D_i)$  between the detection command and the document is defined by the following equations (16) to (18).

$$40 \quad \text{Dist}(Q, D_i) = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{pmatrix} \quad (16)$$

45 and

$$50 \quad z_j = \text{Func}(\text{Dist}'(X_j, Y_j)) \quad (17)$$

where

$$55 \quad \text{Dist}'(X_j, Y_j) = |X_j - Y_j|/M \quad (18)$$

In the above equation (18), a symbol [A] for an arbitrary matrix A represents the meaning defined by the following equation (19).

$$5 \quad |A| = \begin{pmatrix} \sum b_{1m} \\ \sum b_{2m} \\ \sum b_{3m} \\ \vdots \\ \sum b_{Km} \end{pmatrix} \quad (19)$$

10

where, for each element  $a_{ij}$  of the matrix A, if  $a_{ij} < 0$  then  $b_{ij} = -a_{ij}$ , and otherwise  $b_{ij} = a_{ij}$ .

Also, in the above equation (17), the function Func(C) for an arbitrary argument C is defined by the following equation.

$$15 \quad \text{Func}(C) = (\sum_{k=1}^K c_k \times w_k) / \sum_{k=1}^K w_k \quad (20)$$

20

where  $w_k$  is a weight for  $c_k$ , and  $c_k$  is the k-th element of the argument C which can be expressed by the following equation (21).

$$25 \quad C = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_K \end{pmatrix} \quad (21)$$

30

In addition, in this twelfth embodiment, the user can freely register the desired viewpoint extraction rule, and change the setting of the weight factor for each viewpoint.

Namely, the registration of the viewpoint extraction rule and the changing of the weight factor for each viewpoint can be carried out as shown in Figs. 55A to 55D.

Fig. 55A shows a display for entering the character string for the registration of the viewpoint extraction rule, which includes displayed buttons of "register", "viewpoint", and "confirm".

40 The character string entered in the display shown in Fig. 55A is then analyzed at the input analysis unit 202, to obtain the analysis result as the matching portion of the viewpoint extraction rule which is presented as shown in Fig. 55B. Here, the weight is set to an initial setting of 1.

Next, when the user selects the displayed button "viewpoint", a menu enlisting all the available viewpoints appears as shown in Fig. 55C, from which the user can choose any desired viewpoint to be registered in correspondence to the displayed matching portion.

45 Fig. 55D shows a state in which the user has selected the viewpoint of "object" in the menu of Fig. 55C, and then changed the weight to 2 while deleting "improve" from the matching pattern. At this point, when the user selects the displayed button "confirm", the viewpoint extraction rule shown in Fig. 55D is stored in to the viewpoint extraction rule dictionary in the input analysis unit 202 to complete the registration.

50 Next, the thirteenth embodiment of a document detection system according to the present invention will be described in detail.

In this thirteenth embodiment, the first embodiment described above is modified such that the detection record displayed on the window for the detection record display can be edited, and the detection commands and the detection results for a plurality of specified detection records can be presented to facilitate the comparison.

Namely, in this thirteenth embodiment, the original detection record as shown in Fig. 56A can be modified into the modified detection record as shown in Fig. 56B according to the user's input to remove

one unnecessary detection node for example.

In addition, in this thirteenth embodiment, a displayed button "comparison" is provided on the window for the detection record display as shown in Fig. 57, and when this displayed button "comparison" is selected by the user after specifying two detection nodes indicated by square and circular enclosures around the dots representing the detection nodes, the comparison data indicating the number of detected documents, the number of OK documents, and the number of NG documents for each of these detection nodes as well as the number of detected documents, the number of OK documents, and the number of NG documents common to both of these detection nodes are presented as shown in Fig. 57, so as to facilitate the comparison of these two detection nodes by the user.

10 Next, the fourteenth embodiment of a document detection system according to the present invention will be described in detail.

In this fourteenth embodiment, the browsing unit 209 in the first embodiment described above is modified such that document link data indicating relationships between any two documents can be set up on the window for the browsing unit display, and the set up document link data can be indicated on the 15 browsing unit 209.

Namely, in this fourteenth embodiment, the document link data can be expressed in a format of:

`<document ID (1)> ; <link name> ; <document ID (2)>`

where the link name is given by a character string such as "reference" and "original".

In this case, the window for the browsing unit display is given in an exemplary form shown in Fig. 58, 20 on which a displayed button "link" is provided. When the user selects this displayed button "link" after specifying one window for the browsing unit display other than this one, a menu enlisting all the available link names is presented as shown in Fig. 58, from which the user can choose any desired link name appropriate for the intended relationship to be established between the document displayed on this window and the document displayed on the specified window. Then, the browsing unit 209 sets up the document 25 link data in the above described format from the document ID of the document displayed on the specified window, the link name selected from the menu, and the document ID of the document displayed on this window.

Here, the document link data set up by the browsing unit 209 can be indicated on the windows for the 30 browsing unit display as shown in Fig. 59 by connecting the windows displaying the linked documents by a pointer arrow.

In addition, in this fourteenth embodiment, it is also possible to provide an individual data for each document in a data format shown in Fig. 60, in which pointers to document IDs of the OK documents, document IDs of the NG documents, document link data list, and detection node ID list, etc. are collectively registered. Here, it is possible to include the pointers to a part or a whole of the detection records or the 35 detection results in this individual data. In addition, as in the detection node ID list pointed by the pointer to the detection node ID list shown in Fig. 60, it is also possible for the user to add memo to any desired part of the individual data at a time of storing this individual data.

Next, the fifteenth embodiment of a document detection system according to the present invention will be described in detail.

40 In this fifteenth embodiment, the detection result display in the document display mode by the detection result display unit 207 in the sixth embodiment described above is modified such that the positioning of the axes and the assignment of the viewpoint to each axis in the multi-dimensional display can be specified by the user.

Namely, in this fifteenth embodiment, the window for the detection result display in the document 45 display mode is given as shown in Fig. 61 which includes a displayed button "positioning" instead of the display button "arrange" in Fig. 42 for the sixth embodiment.

When the user selected this displayed button "positioning", a menu enlisting possible positioning patterns is presented as shown in Fig. 61, from which the user can choose any desired positioning pattern item to change the positioning of the axes in the detection result display, where "A", "B", "C" used in the 50 menu represents the assigned viewpoints as indicated in the detection result display. This menu also includes items for the one-dimensional display and the two-dimensional display, in response to which the dimensionality of the detection result display can be changed accordingly.

In addition, on the window for the detection result display in the document display mode of this fifteenth embodiment, the labels indicating the viewpoints assigned to the axes of the detection result display are 55 also made to be displayed buttons that can be selected by the user. For example, when the user selects the displayed button "conclusion" assigned to one axis as shown in Fig. 62, a menu enlisting the other available viewpoints is presented as shown in Fig. 62, from which the user can choose any desired viewpoint to be assigned to this axis instead of the current assignment of "conclusion".

Furthermore, in this fifteenth embodiment, the adaptive learning function can be provided for the purpose of learning the viewpoint assignments and the axes positioning selected by each user, such that the same viewpoint assignments and the axes positioning preferred by each user can be resumed at a time of the subsequent detection operation by the same user automatically.

5 Next, the sixteenth embodiment of a document detection system according to the present invention will be described in detail.

In this sixteenth embodiment, the sixth embodiment described above is modified such that the display of the windows for the browsing unit display generated during the previous detection operations can be continued during the subsequent detection operation, while distinguishing the windows belonging to 10 different detection operations by changing the window display manners.

Namely, in this sixteenth embodiment, the windows for the browsing unit display generated during the previous detection operations and the windows for the browsing unit display generated during the subsequent detection operation are distinguished by changing the colors or the concentration levels of the frame or background of each window according to the order of detection operations for example.

15 Fig. 76 shows an exemplary state of the display of the windows for the browsing unit display in which the windows having the blackened side and lower edges belong to the previous detection operation, in contrast to the other windows belonging to the present detection operation.

It is to be noted here that, besides those already mentioned above, many modifications and variations 20 of the above embodiments may be made without departing from the novel and advantageous features of the present invention. Accordingly, all such modifications and variations are intended to be included within the scope of the appended claims.

### Claims

- 25 1. A document detection system, comprising:
  - document memory means for storing a plurality of documents;
  - input means for entering user's input for commanding a document detection in the documents stored in the document memory means;
  - input analysis means for analyzing the user's input entered by the input means to extract keywords and viewpoints relevant to each keyword contained in the user's input, and constructing a detection command from the keywords and the viewpoints extracted from the user's input; and
  - detection means for detecting those documents stored in the document memory means which match with the detection command constructed by the input analysis means as detected documents of a detection result.
- 35 2. The system of claim 1, wherein the input analysis means extracts the viewpoints relevant to each keyword according to prescribed viewpoint extraction rules specifying expression patterns to be matched with a part of the user's input containing each keyword, and the viewpoints corresponding to the expression patterns.
- 40 3. The system of claim 1, wherein the detection means detects the detected documents by searching a set of each keyword and the viewpoint relevant to said each keyword through a keyword index enlisting all the keywords of the documents stored in the document memory means in relation to the viewpoints.
- 45 4. The system of claim 1, further comprising:
  - distance calculation means for calculating distances of the detected documents detected by the detection means with respect to the detection command, for each viewpoint; and
  - detection result display means for presenting a detection result display indicating the detection result obtained by the detection means, in a multi-dimensional display formed by setting the viewpoints to axes with the detection command as an origin and using the distances of the detected documents for each viewpoint calculated by the distance calculation means as coordinates of the detected documents with respect to each axis representing each viewpoint.
- 55 5. The system of claim 4, further comprising:
  - record management means for managing a pair of the detection command and the detection result as a detection record; and
  - detection record display means for presenting a detection record display indicating relations among a plurality of detection records managed by the detection record management means in a form of a

tree structure in which each detection record is represented by a node in the tree structure.

6. The system of claim 5, wherein the detection record display means commands the detection result display means to present the detection result display for the detection record represented by the node specified by a user on the detection record display.
- 5
7. The system of claim 4, further comprising:  
means for entering accept/reject data for indicating an acceptance/rejection of each detected document in the detection result display, wherein the detection result display means indicates those detection documents which are accepted by the accept/reject data distinguishably from those detection documents which are rejected by the accept/reject data.
- 10
8. The system of claim 7, wherein the detection means constructs a new detection command according to the accept/reject data, and re-detects those documents stored in the document memory means which match with the new detection command constructed.
- 75
9. The system of claim 4, further comprising:  
browsing means for displaying data indicative of each detected document specified by a user on the detection result display, wherein the detection result display means indicates each detected document for which the browsing means displays the data distinguishably from other detected documents on the detection result display.
- 20
10. The system of claim 9, wherein the browsing means displays at least one of a summary, a keyword list, and a viewpoint list of each detected document as the data indicative of each detected document.
- 25
11. The system of claim 9, wherein the browsing means generates a window for separately displaying the data for each detected document specified by the user on the detection result display, and arranges a plurality of windows for displaying the data for a plurality of the detected documents specified by the user in a desired order specified by the user.
- 30
12. The system of claim 4, wherein the detection result display means rotates the detection result display according to an input for specifying a rotation entered by a user on the detection result display.
13. The system of claim 4, wherein the detection result display means enlarges/contracts the detection result display according to an input for specifying a range in the multi-dimensional display entered by a user on the detection result display.
- 35
14. The system of claim 4, wherein the detection result display means presents the detection result display in a cluster display mode in which a multi-dimensional space of the multi-dimensional display is divided into a number of sub-spaces, and the detected documents located within each sub-space is represented collectively as a cluster.
- 40
15. The system of claim 14, wherein the detection result display means also presents the detection result display in a document display mode which includes the multi-dimensional display for only a cluster specified by a user on the detection result display in the cluster display mode.
- 45
16. The system of claim 14, wherein the detection result display means changes the detection result display in the cluster display mode according to a number of the detected documents contained in each cluster.
- 50
17. The system of claim 14, wherein the detection result display means also enters accept/reject data for indicating an acceptance/rejection of all the detected documents in each cluster in the detection result display in the cluster display mode collectively.
- 55
18. A method of document detection, comprising the steps of:  
analyzing a user's input for commanding a document detection to extract keywords and viewpoints relevant to each keyword contained in the user's input;  
constructing a detection command from the keywords and the viewpoints extracted from the user's

input at the analyzing step; and

detecting those documents among a plurality of stored documents which match with the detection command constructed at the constructing step as detected documents of a detection result.

5   **19.** The method of claim 18, further comprising the steps of:

calculating distances of the detected documents obtained at the executing step with respect to the detection command, for each viewpoint; and

10   presenting a detection result display indicating the detection result obtained at the executing step, in a multi-dimensional display formed by setting the viewpoints to axes with the detection command as an origin and using the distances of the detected commands for each viewpoint calculated at the calculating step as coordinates of the detected documents with respect to each axis representing each viewpoint.

15   **20.** The method of claim 19, wherein the presenting step presents the detection result display in a cluster display mode in which a multi-dimensional space of the multi-dimensional display is divided into a number of sub-spaces, and the detected documents located within each sub-space is represented collectively as a cluster.

20

25

30

35

40

45

50

55

FIG. 1

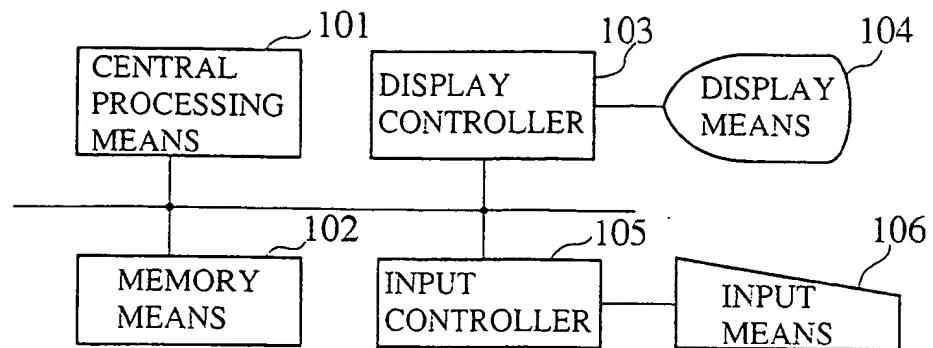


FIG. 2

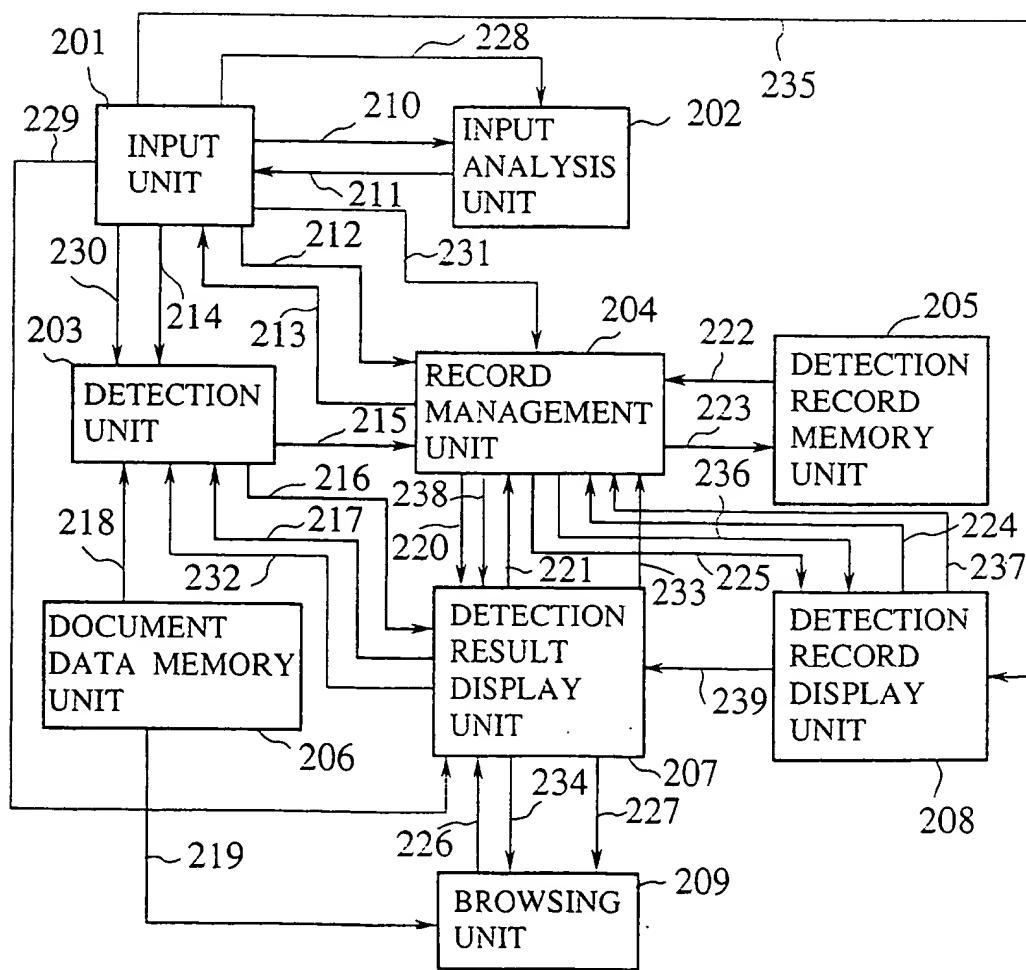


FIG. 3

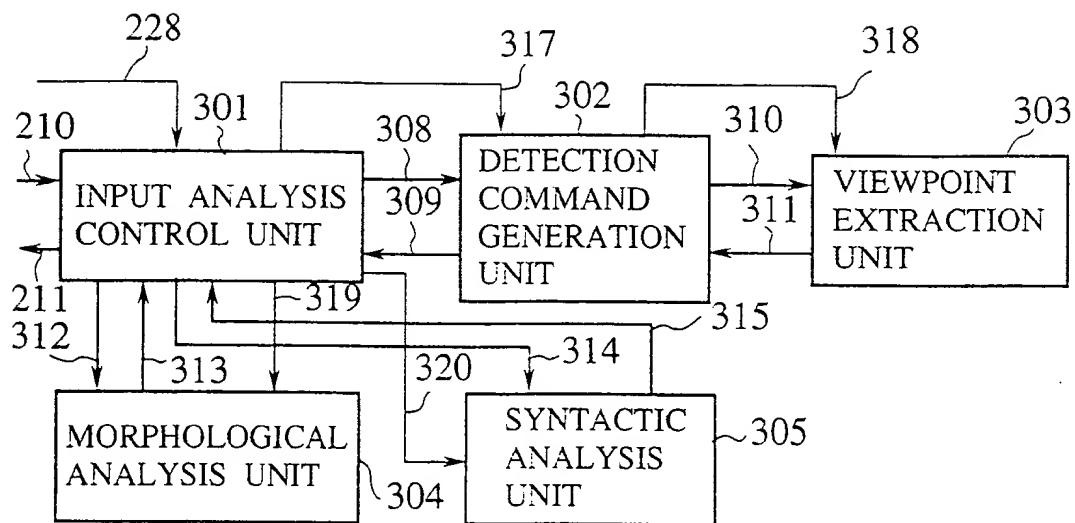


FIG. 4

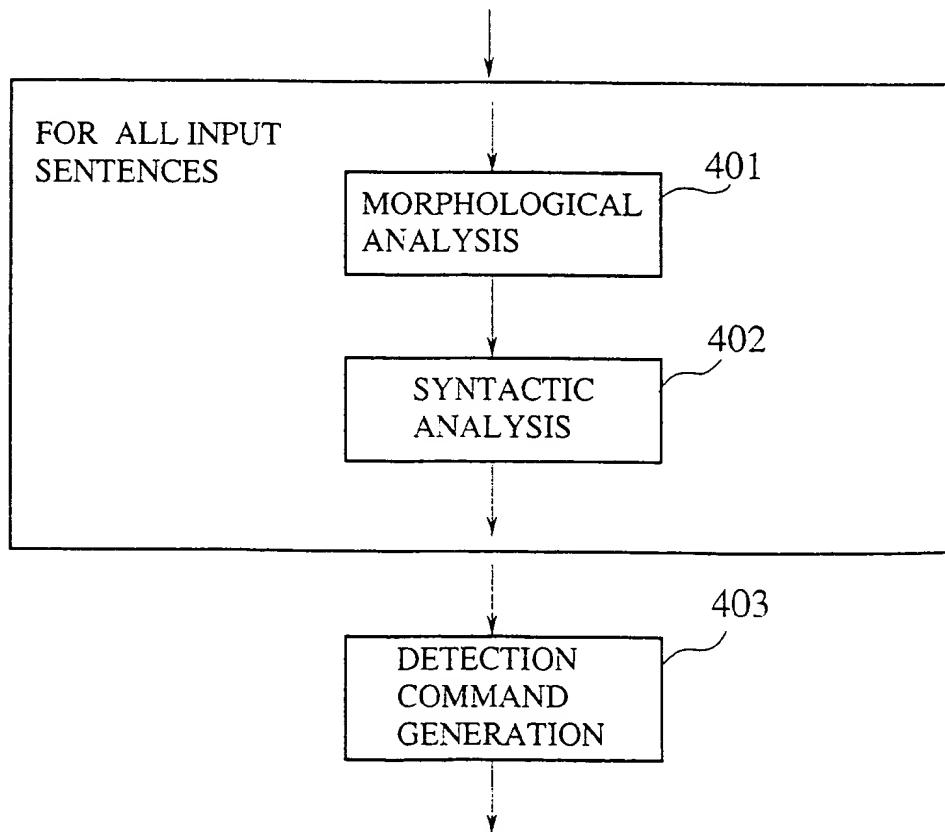


FIG. 5

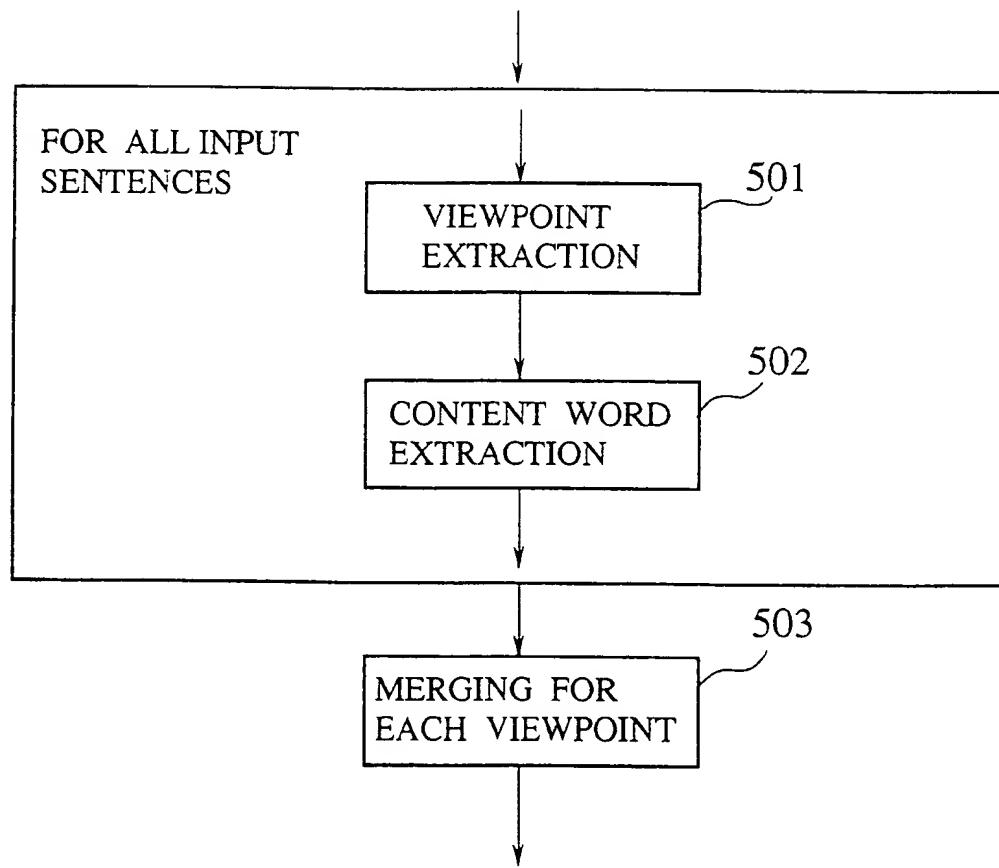


FIG. 6

(WITH AN OBJECT OF)	⇒OBJECT
(CONCLUDE [PAST TENSE] )	⇒CONCLUSION
(OBTAIN [PAST TENSE, PASSIVE VOICE] )	⇒CONCLUSION
(DEVELOP [PAST TENSE] )	⇒CONCLUSION

FIG. 7

(FORMAT)	(EXAMPLES)
<KEYWORD> :	COMPUTER : OBJECT
<VIEWPOINT LIST>	DESIGN : OBJECT, CONCLUSION
	TOOL : CONCLUSION
	EXAMPLE : CONCLUSION

FIG. 8

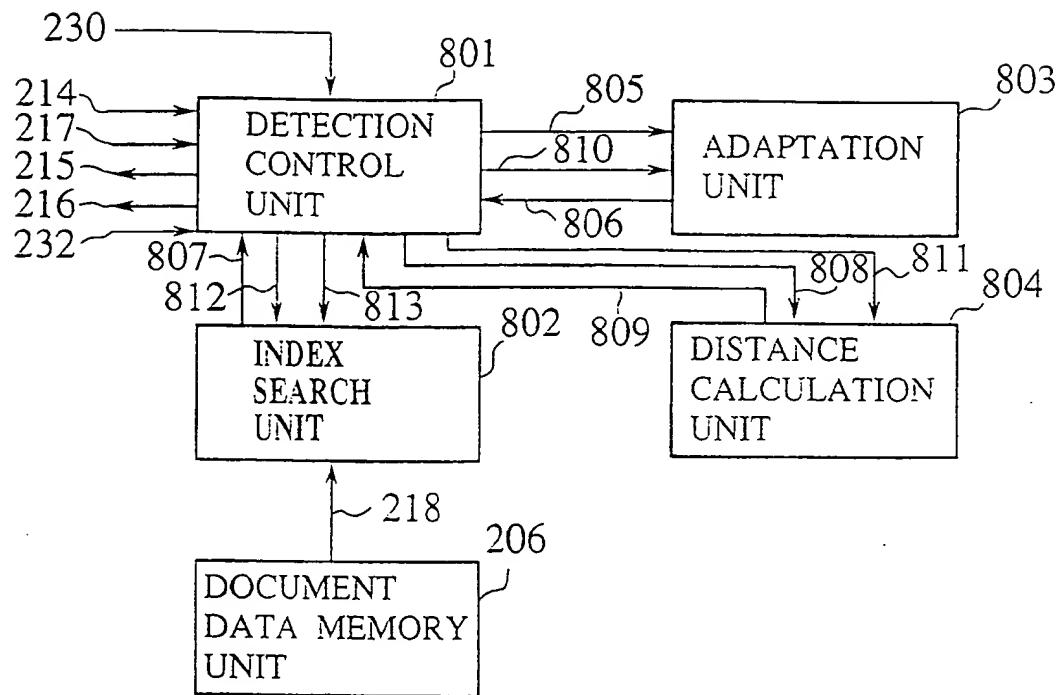


FIG. 9

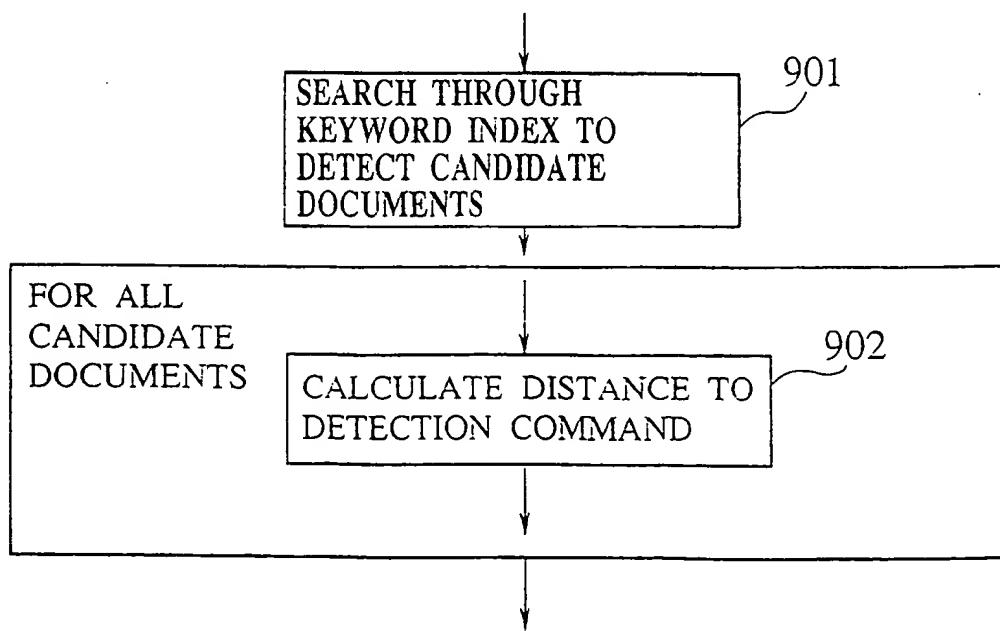


FIG.10

HEAD CHARACTER STORAGE REGION		SUBSEQUENT CHARACTER STORAGE REGION
00053	{ 機, 00305, 0, 0	00305 械, 00623, 0, file4, CONCLUSION, 0
00091	{ 計, 00356, 0, 0	00321 機, 0, file30, OBJECT, CONCLUSION, 0
000AD	{ 算, 00A15, 00A16, 0, 0	00335 計, 0, file30, CONCLUSION, file 31, OBJECT, 0
000B3	{ 事, 00390, 0, 0	00356 算, 00321, 0, file81, BACKGROUND, 0
000B4	{ 自, 00411, 0, 0	00390 象, 0, file45, OBJECT, CONCLUSION, 0
000G1	{ 設, 00335, 0, 0	00411 動, 00623, 0, file21, BACKGROUND, 0
000E5	{ 翻, 00512, 0, 0	00512 訳, 0, file21, OBJECT, 0
		00623 翻, 00914, 0, 0
		00914 訳, 0, file25, OBJECT, file21, CONCLUSION, 0
		00A15 法, 0, file24, OBJECT, BACKGROUND, 0
		00A16 術, 0, file28, CONCLUSION, 0

## FIG.11

### (FORMAT)

```
<DOCUMENT ID1> ; <DISTANCE11> , <DISTANCE12> ; ··; <DISTANCE1N> /  
<DOCUMENT ID2> ; <DISTANCE21> , <DISTANCE22> ; ··; <DISTANCE2N> /  
:  
<DOCUMENT IDK> ; <DISTANCEK1> , <DISTANCEK2> ; ··; <DISTANCEKN> /
```

### (EXAMPLES)

```
file21 ; 0,7,0,7,0,1/  
file28 ; 0,5,0,0,0,9/  
file29 ; 0,4,0,2,0,4/  
file30 ; 0,4,0,1,0,3/  
file35 ; 0,2,0,2,0,5
```

FIG.12

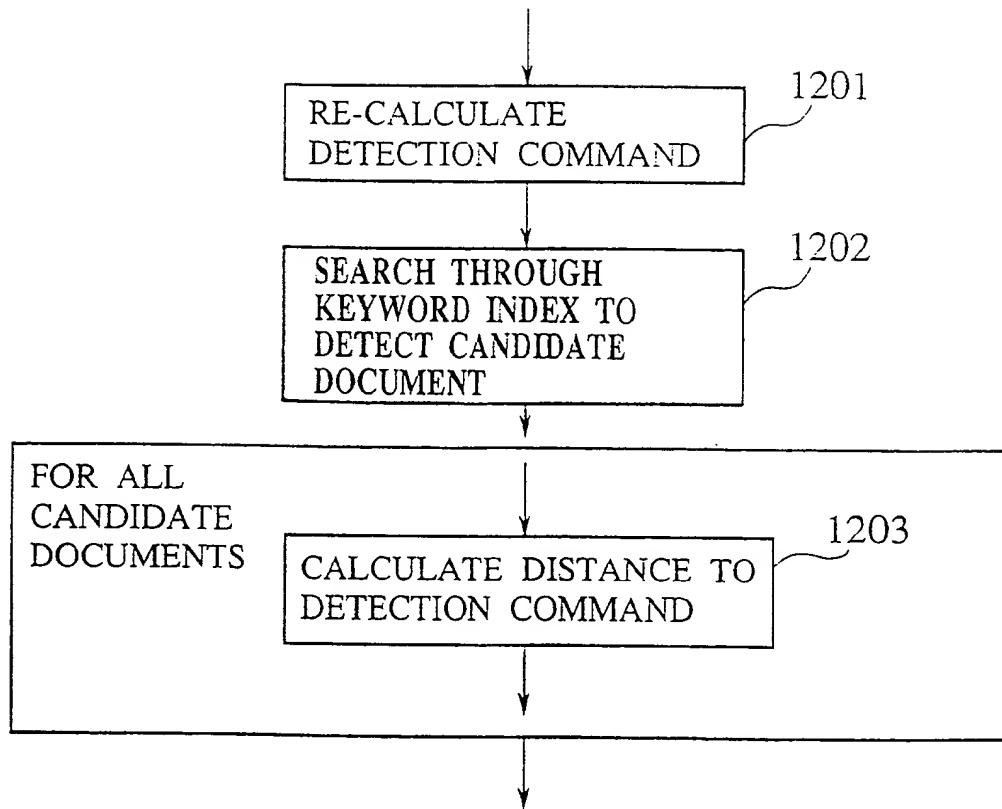


FIG.13

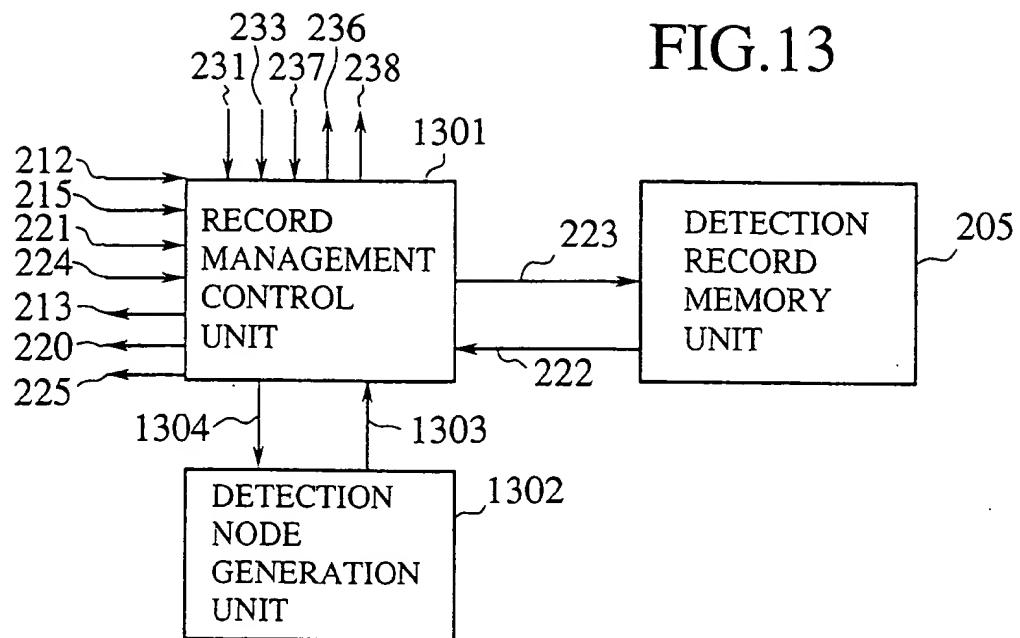


FIG.14

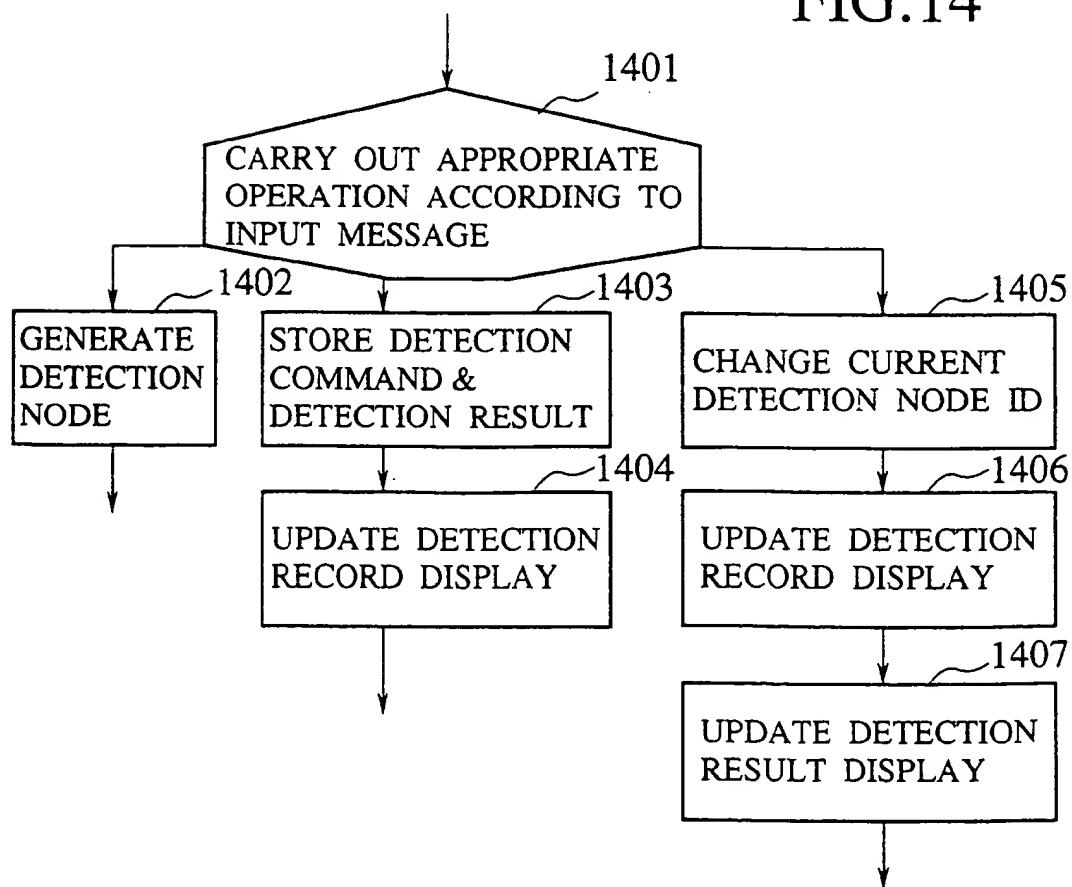


FIG.15

(FORMAT)	(EXAMPLE)
⟨DETECTION NODE ID⟩ .	11.
⟨PARENT NODE ID⟩ .	10.
⟨DETECTION COMMAND⟩ .	COMPUTER : OBJECT / DESIGN : OBJECT,CONCLUSION / TOOL : CONCLUSION
⟨DETECTION RESULT⟩ .	file21 ; 0.7, 0.7, 0.1 / file28 ; 0.5, 0.0, 0.9/ file29 ; 0.4, 0.2, 0.4 / file30 ; 0.4, 0.1, 0.3/ file35 ; 0.2, 0.2, 0.5

FIG.16

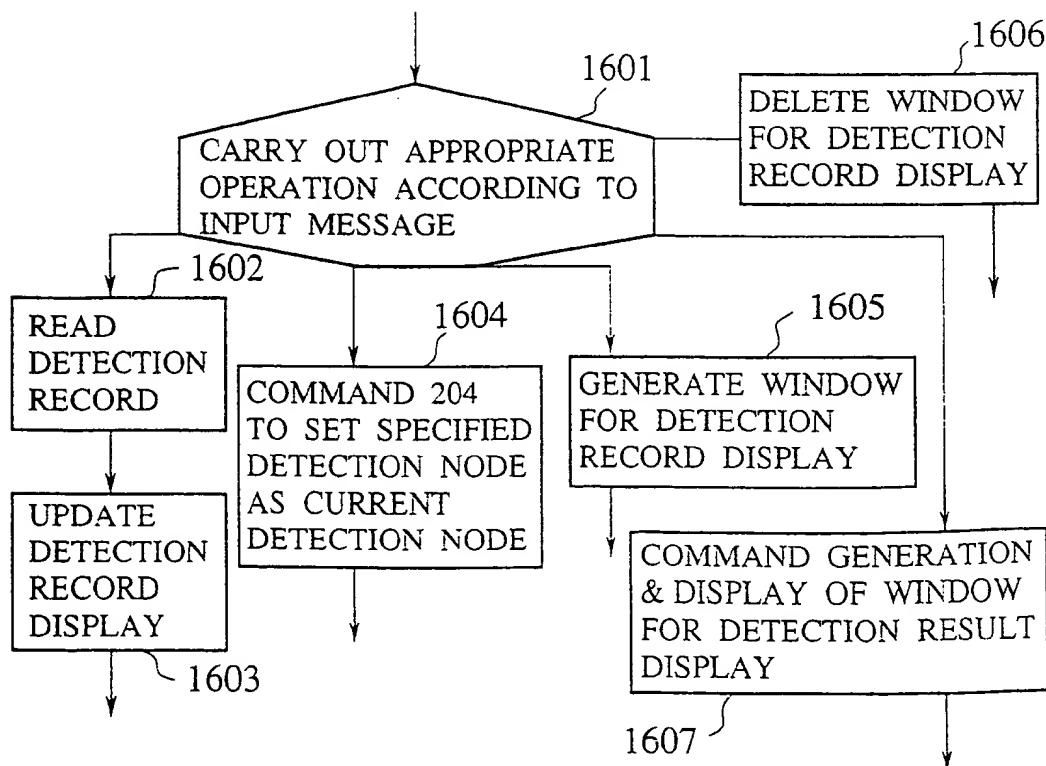


FIG.17A

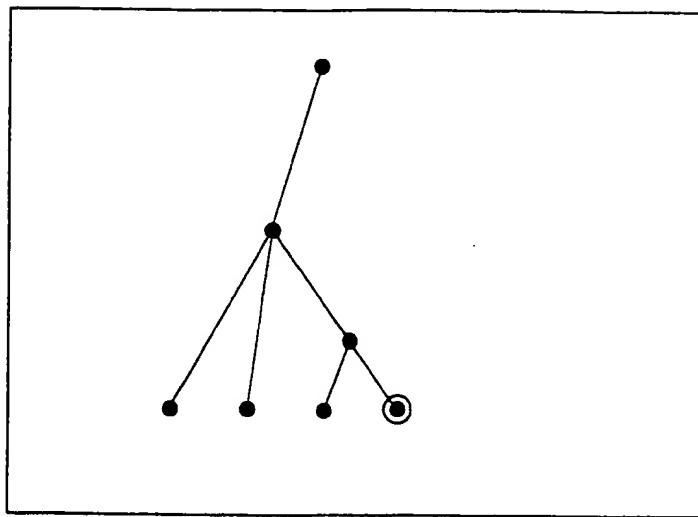


FIG.17B

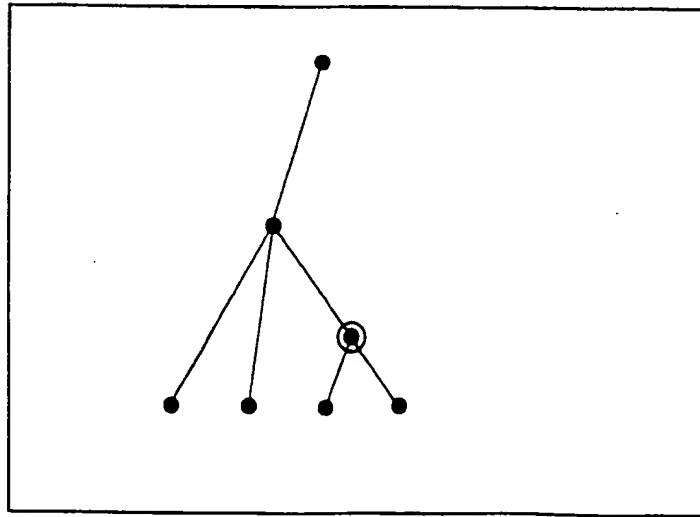
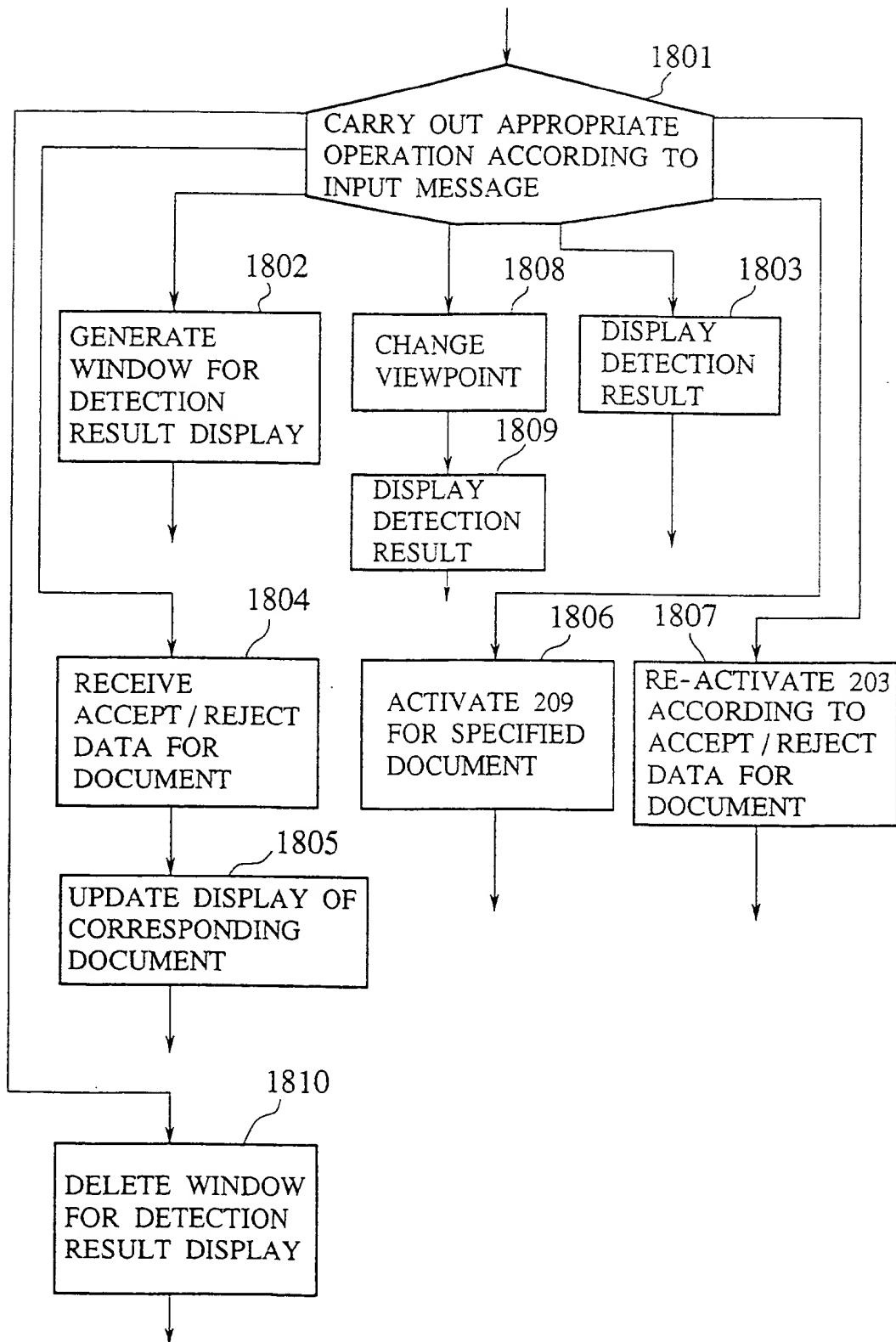
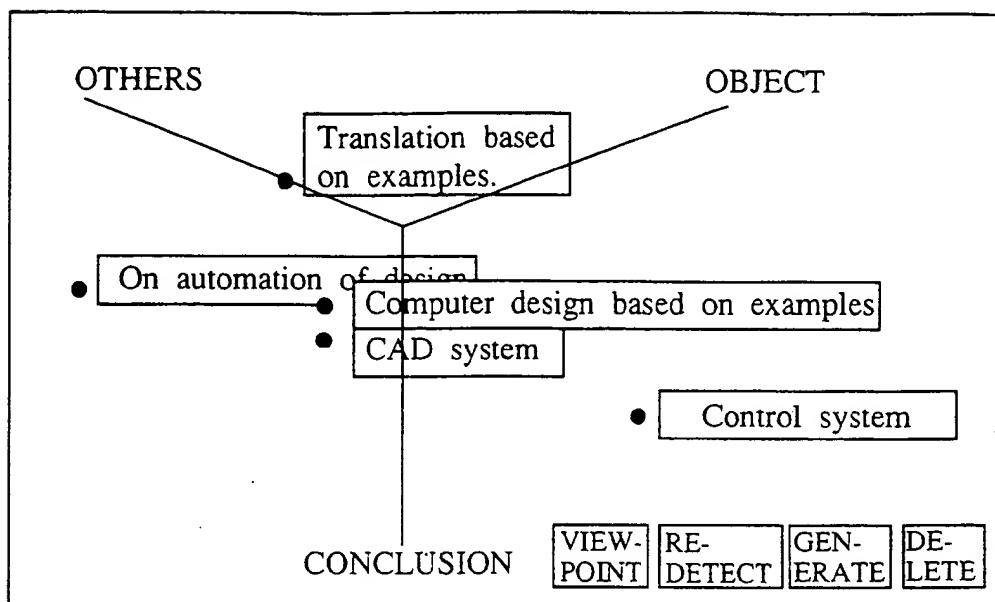


FIG.18

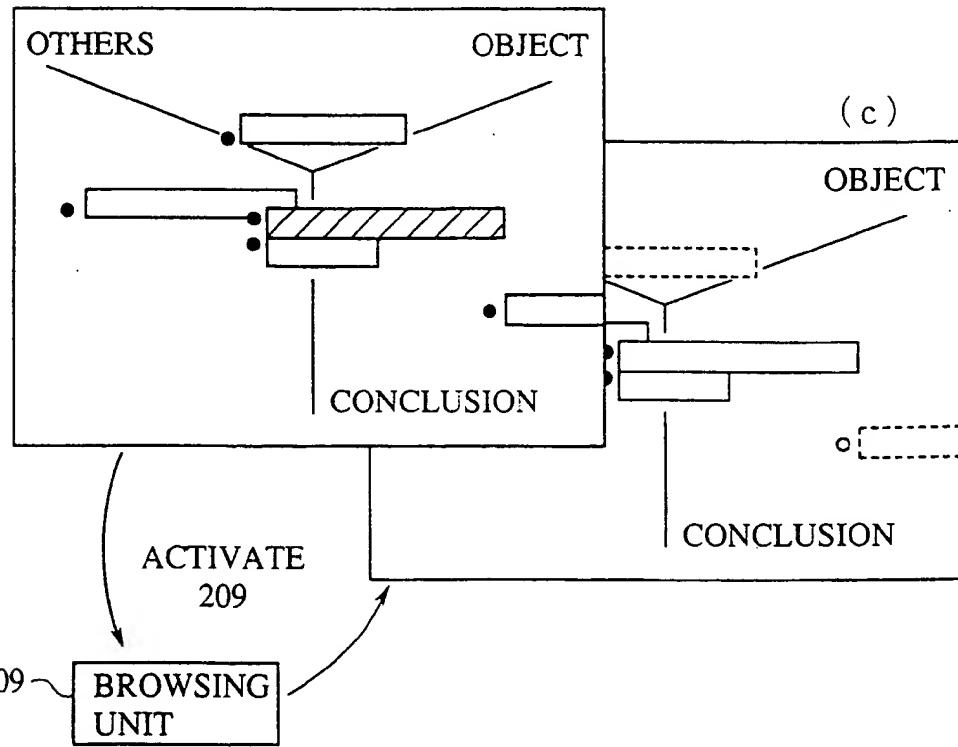


(a)

FIG.19



(b)



(c)

## FIG.20

### (FORMAT)

```
<DOCUMENT ID1> ; <ACCEPT / REJECT DATA 1> /  
<DOCUMENT ID2> ; <ACCEPT / REJECT DATA 2> /  
⋮  
<DOCUMENT IDK> ; <ACCEPT / REJECT DATA K> .
```

### (EXAMPLES)

file 21 ; 0.

file 28 ; 0.

file 30 ; 1.

FIG.21

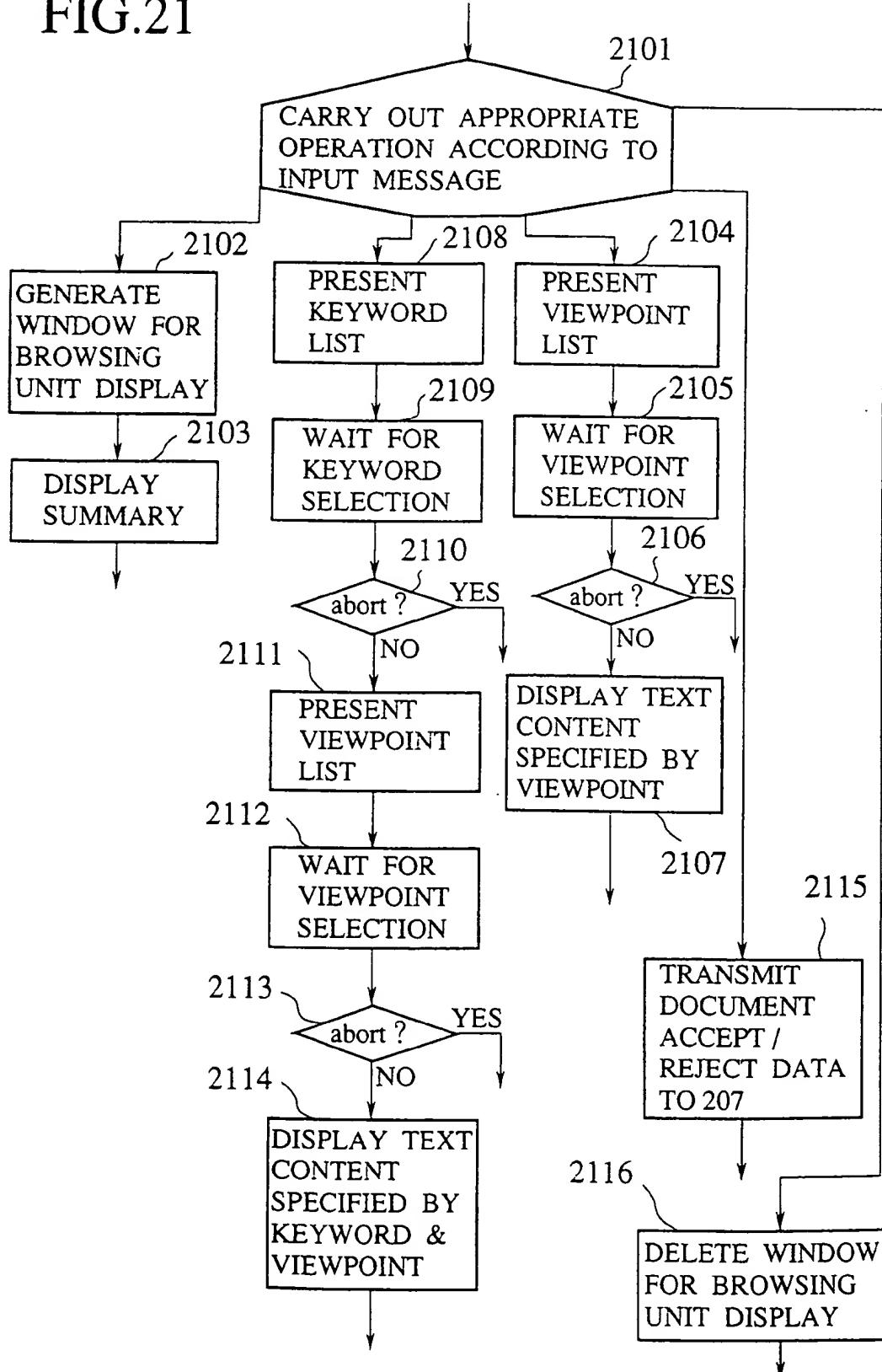


FIG.22

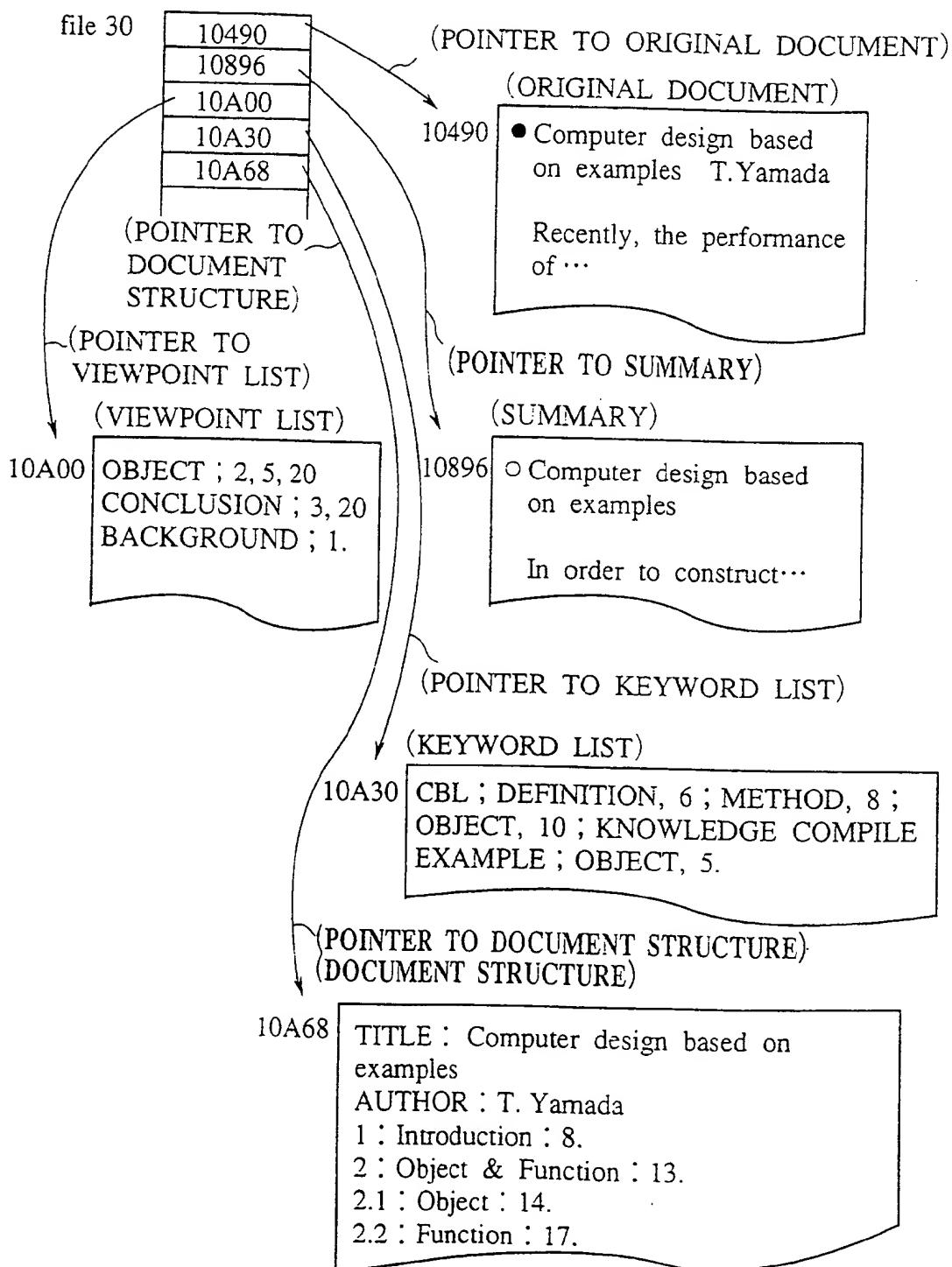


FIG.23

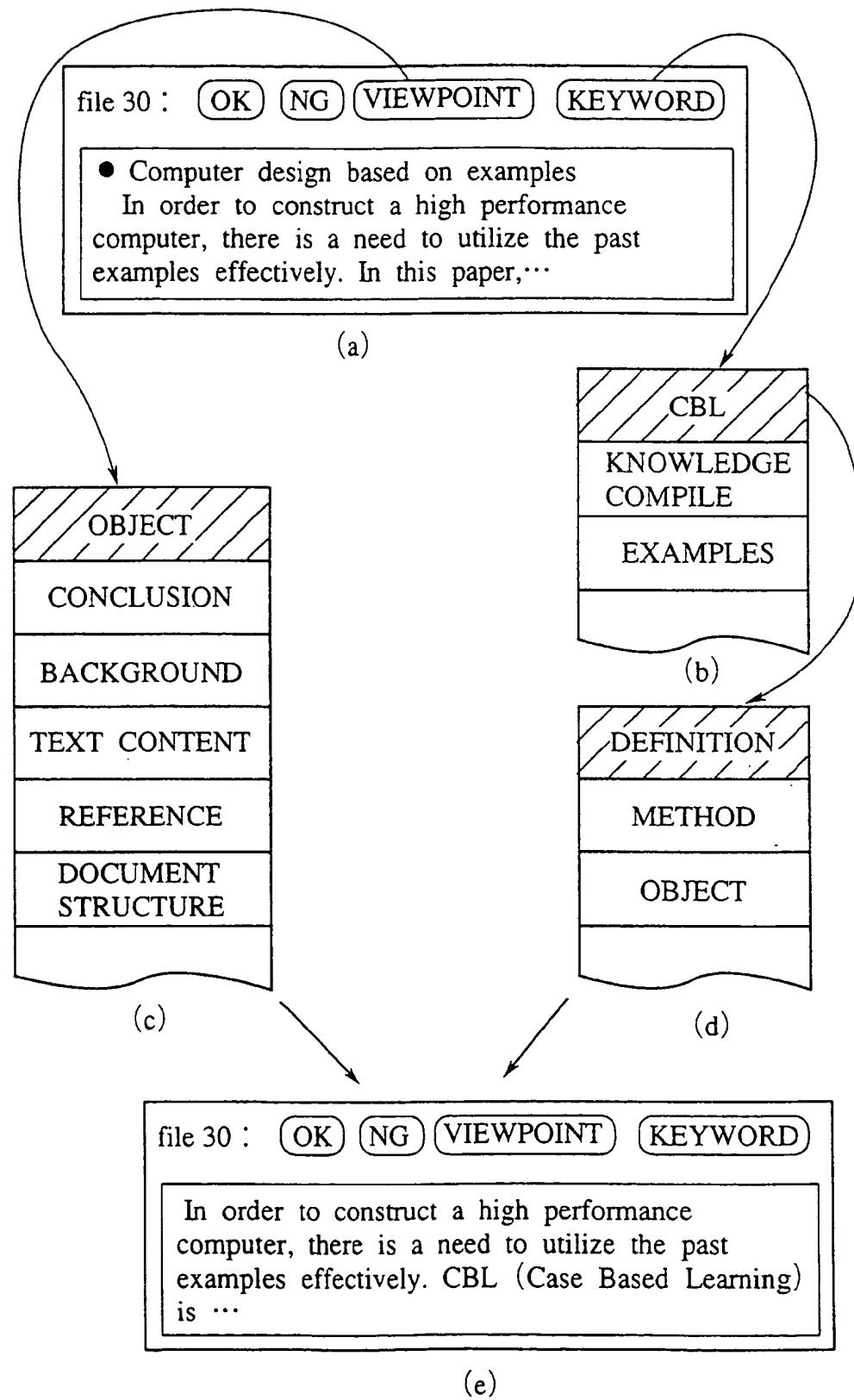


FIG.24A

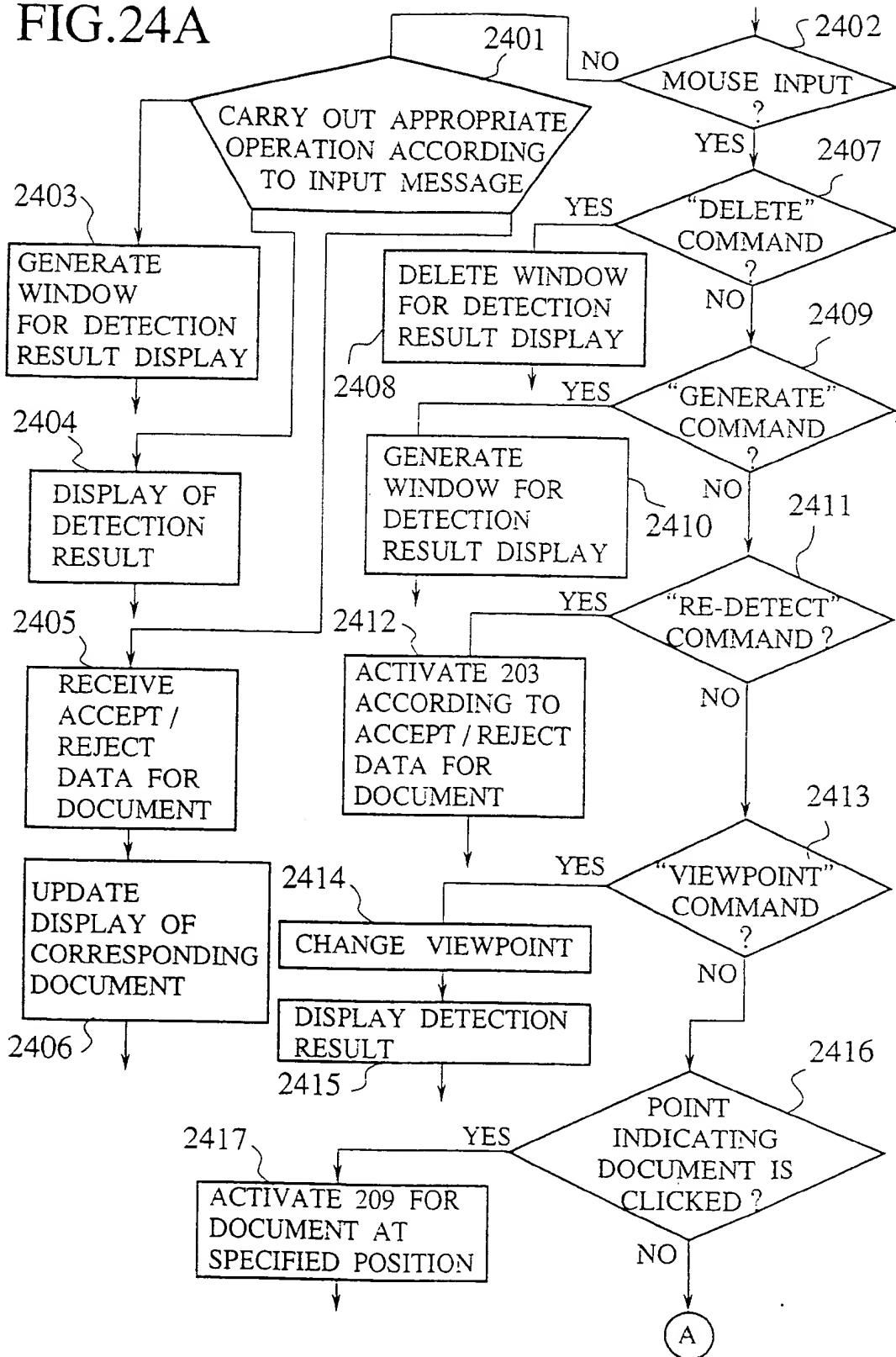


FIG.24B

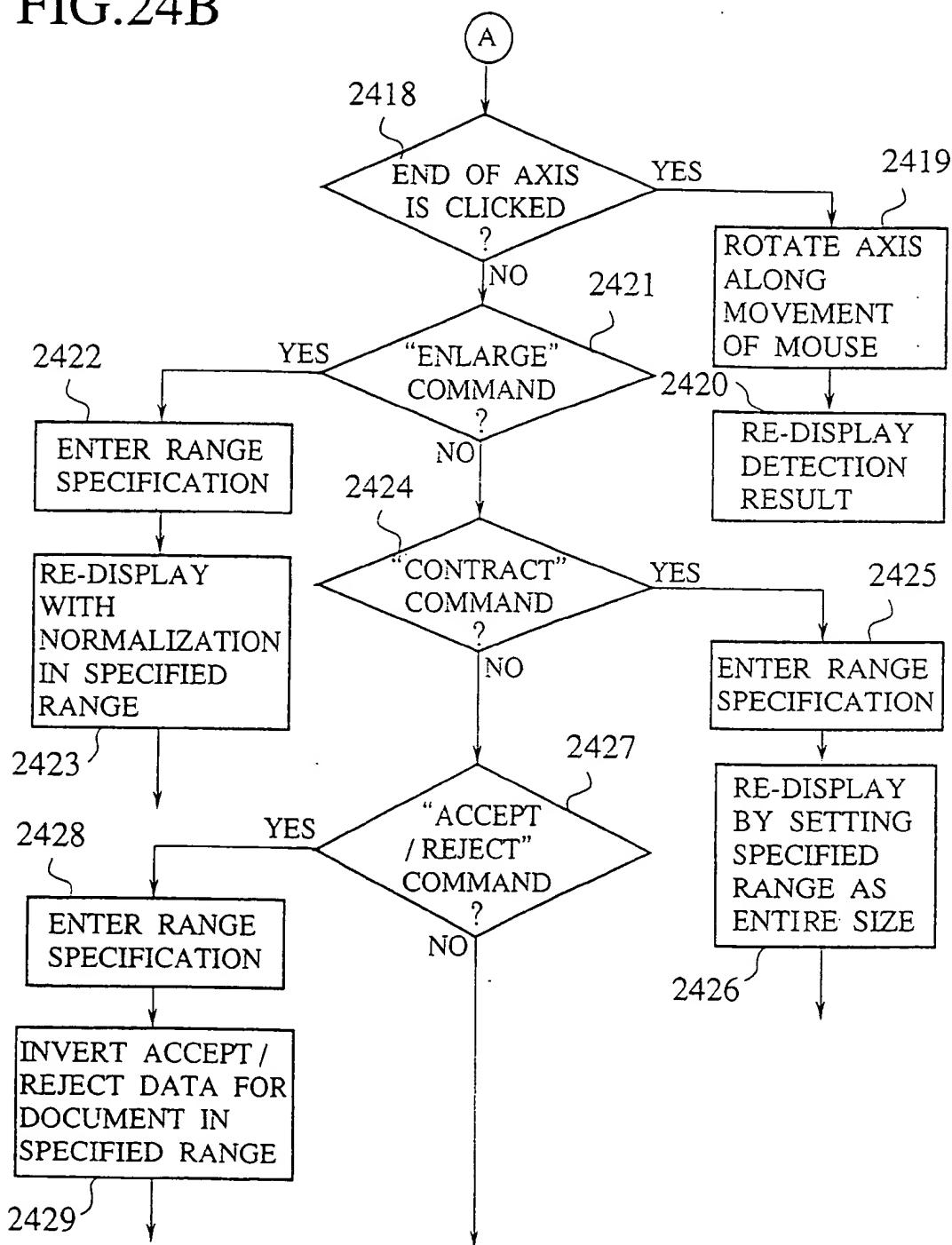
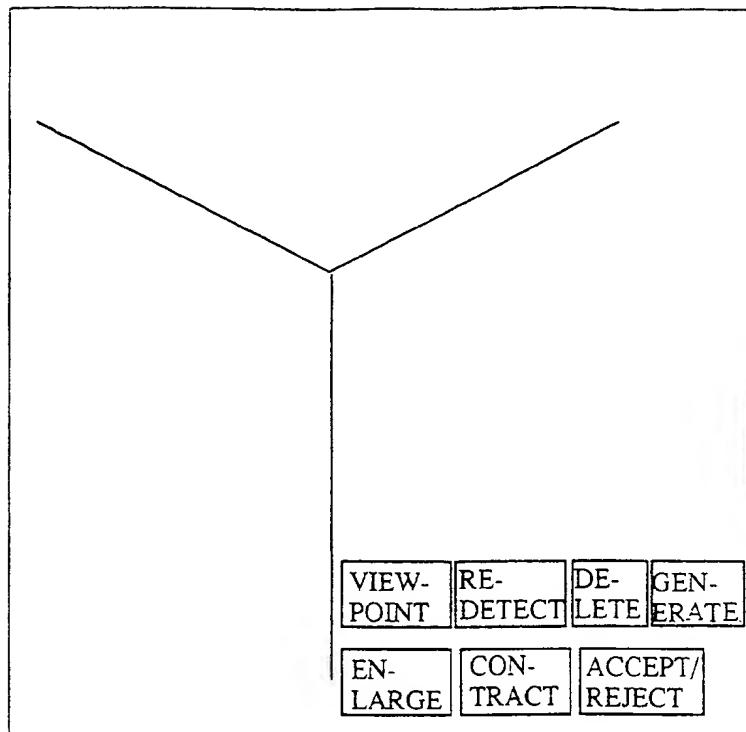
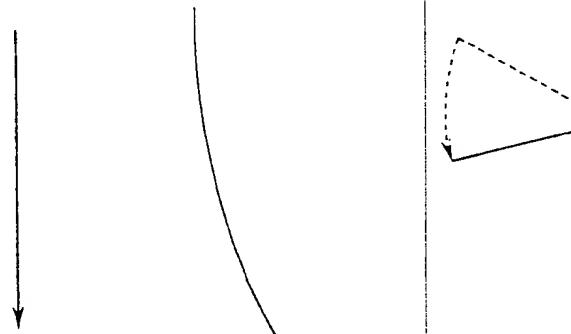


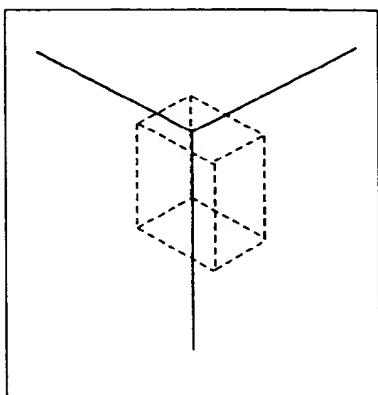
FIG.25



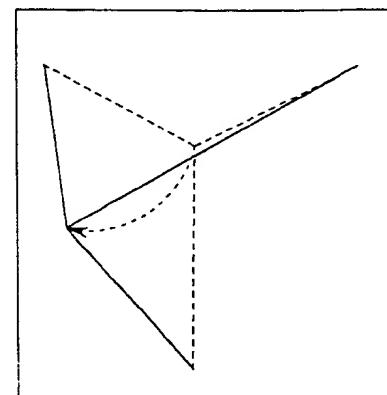
( a )



( b )



( d )



( c )

FIG.26A

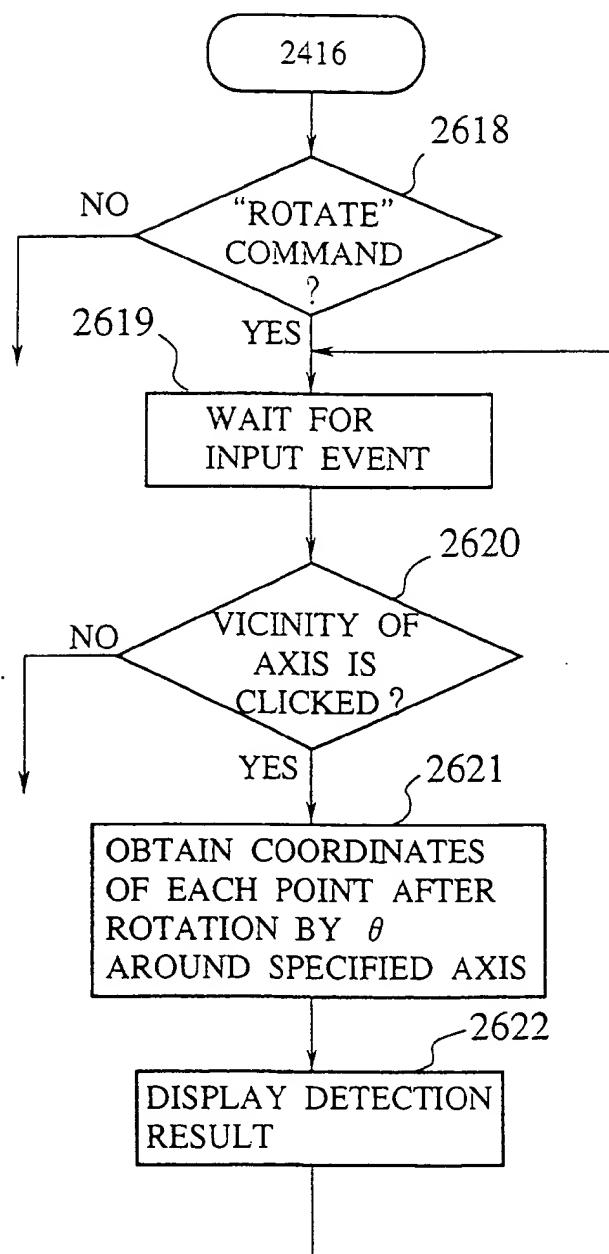


FIG.26B

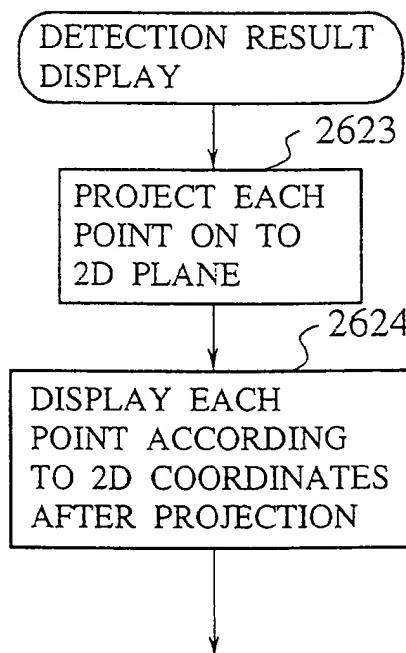


FIG.27A

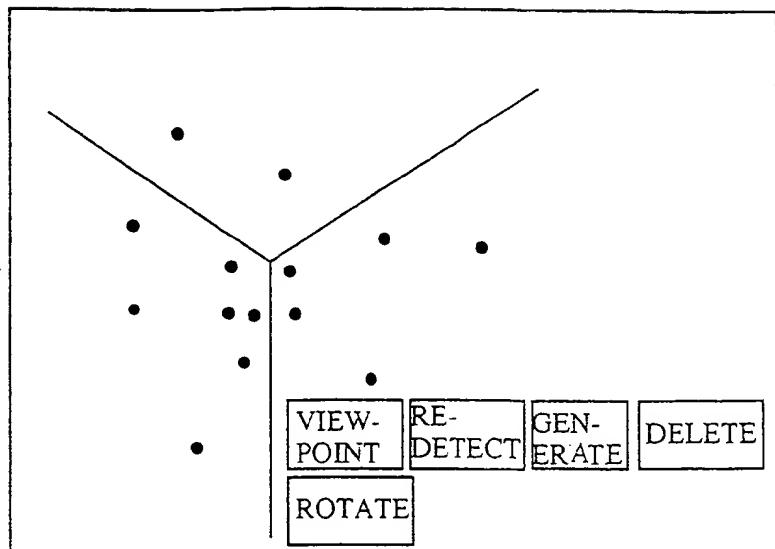


FIG.27B

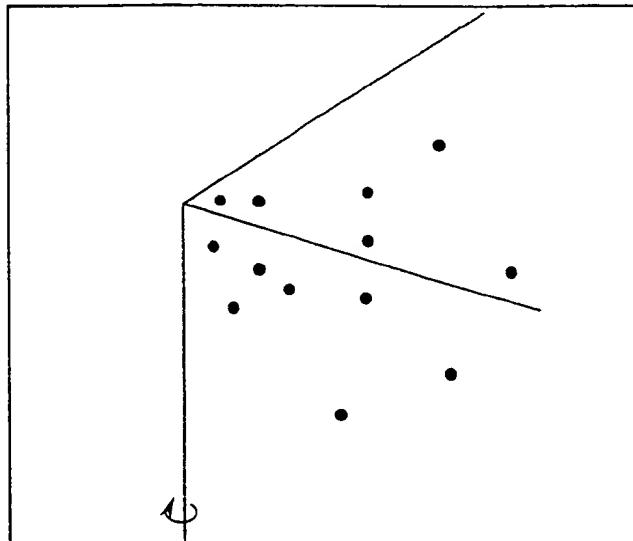


FIG.27C

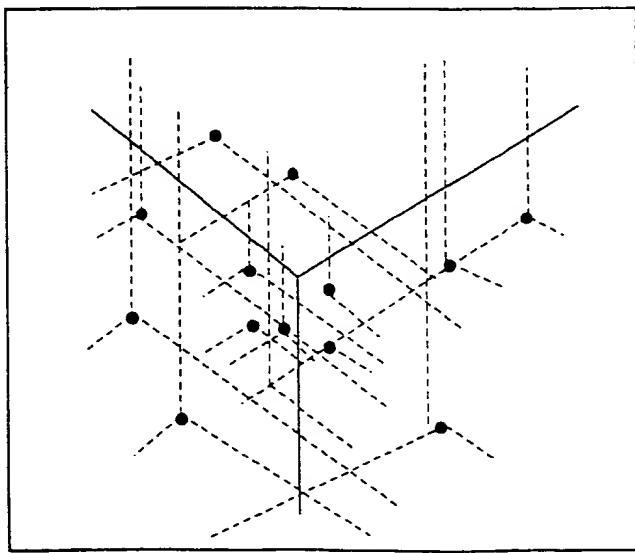


FIG.28

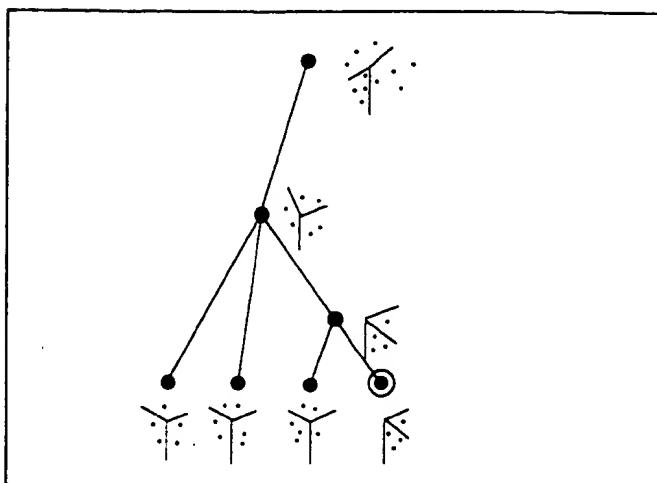


FIG.29

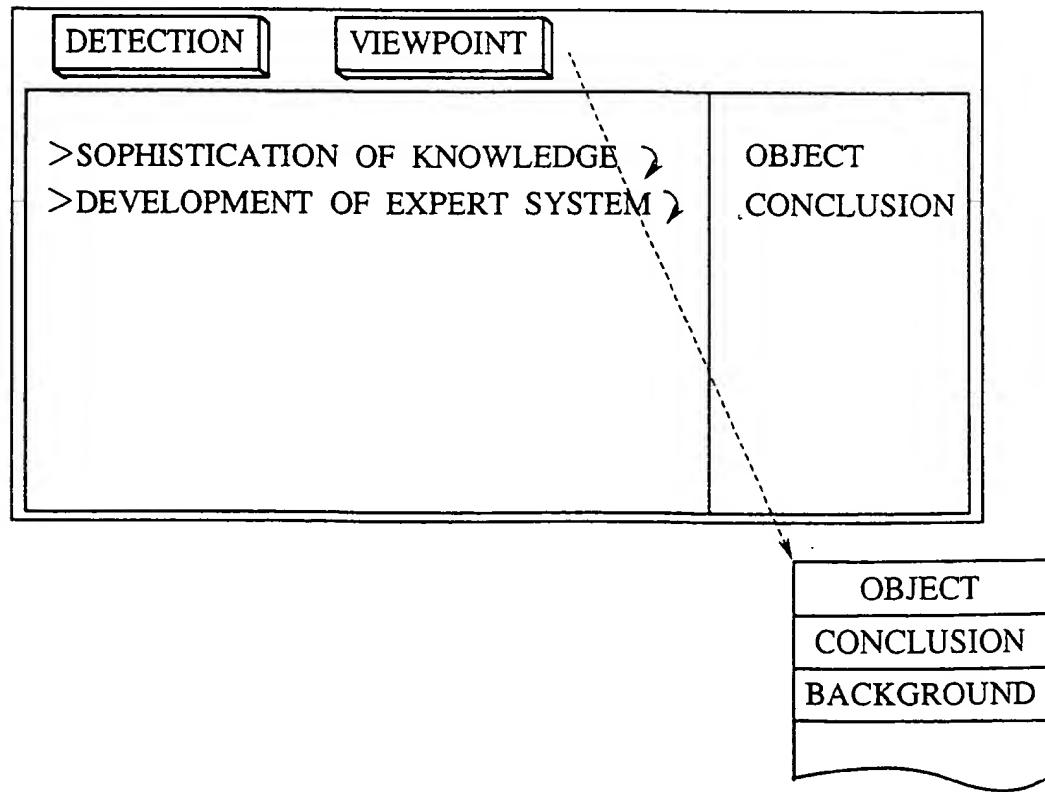


FIG.30

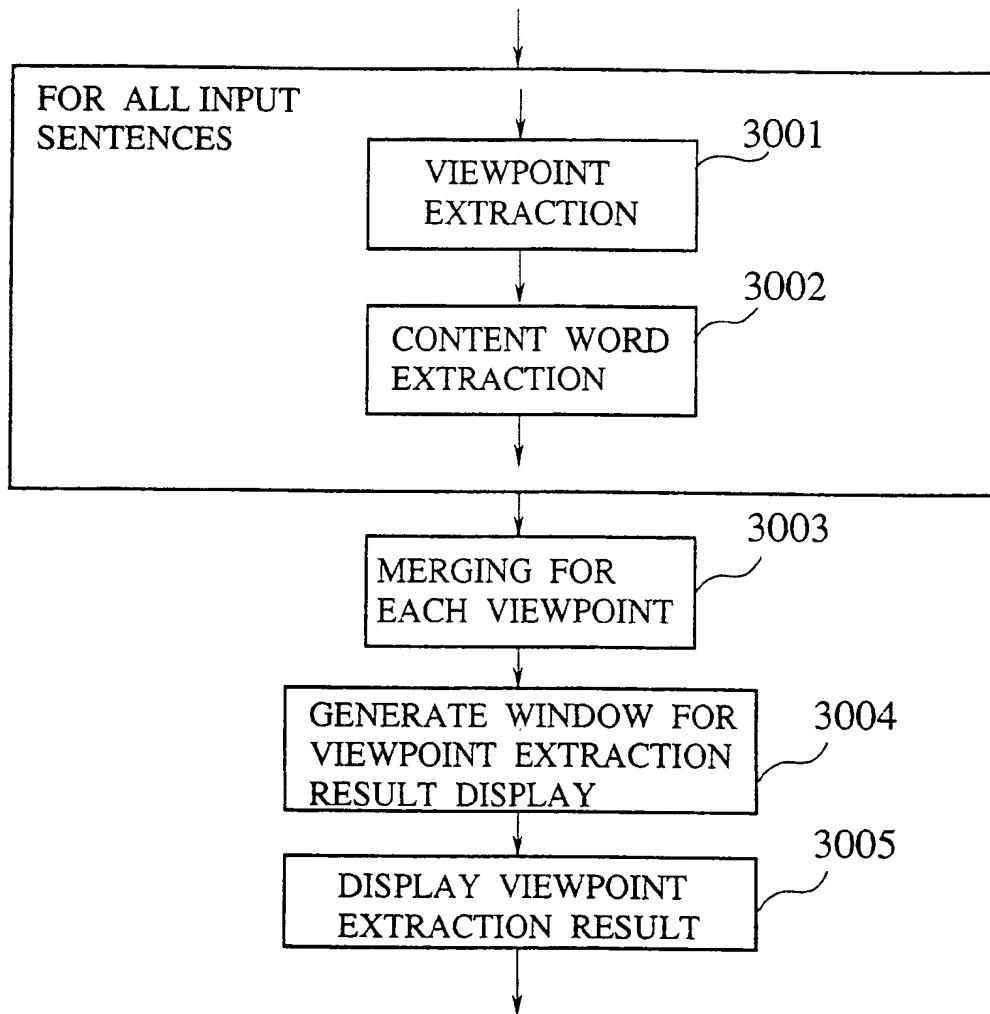


FIG.31

VIEWPOINT	KEYWORD
OBJECT	COMPUTER * DESIGN
CONCLUSION	EXAMPLE * DESIGN * TOOL

FIG.32

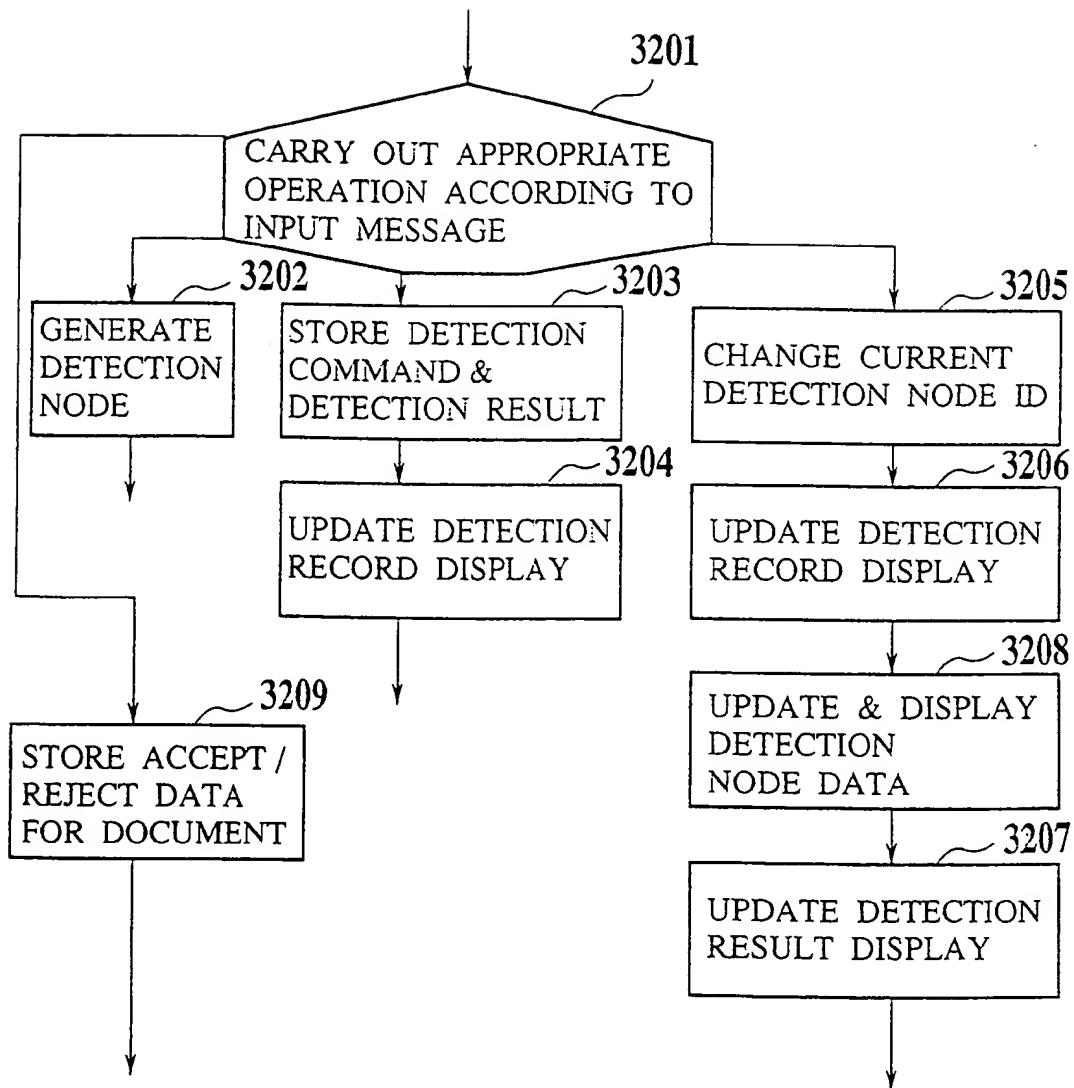


FIG.33

(FORMAT)	(EXAMPLE)
<DETECTION NODE ID> .	12.
<PARENT NODE ID> .	10.
<DETECTION DATE & TIME> .	1993012535846
<DETECTION INPUT RESULT> .	A design tool based on examples developed with an object of a computer design.
<DETECTION COMMAND> .	COMPUTER ; OBJECT/DESIGN ; OBJECT, CONCLUSION/TOOL ; CONCLUSION/EXAMPLE ; CONCLUSION
<# OF DETECTED DOCUMENTS> .	61.
<DETECTION RESULT> .	file21 ; 35, 35, 5 / file28 ; 25, 0, 45/ file29 ; 20, 10, 20 / file30 ; 20, 5, 15/ file35 ; 10, 10, 25
<# OF OK DOCUMENTS> .	1.
< OK DOCUMENTS> .	file 30.
<# OF NG DOCUMENTS> .	2.
< NG DOCUMENTS> .	file21. file28.

FIG.34

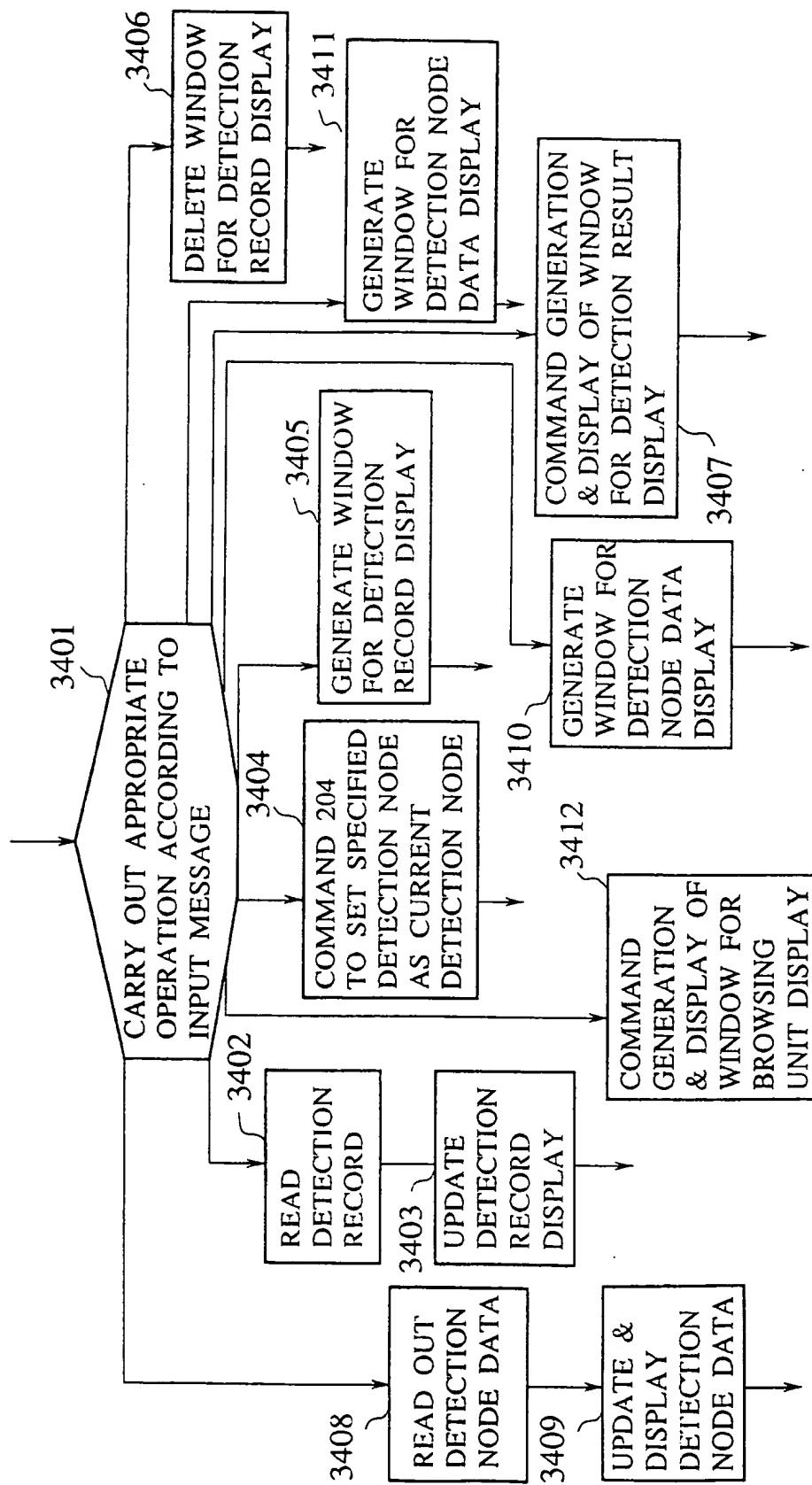


FIG.35A

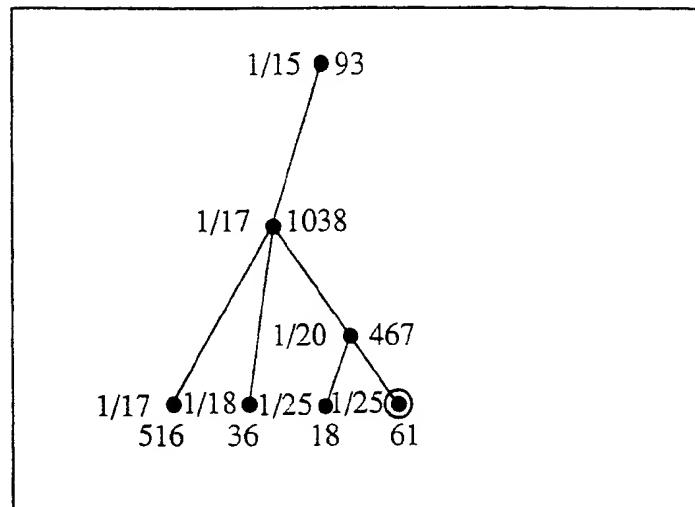
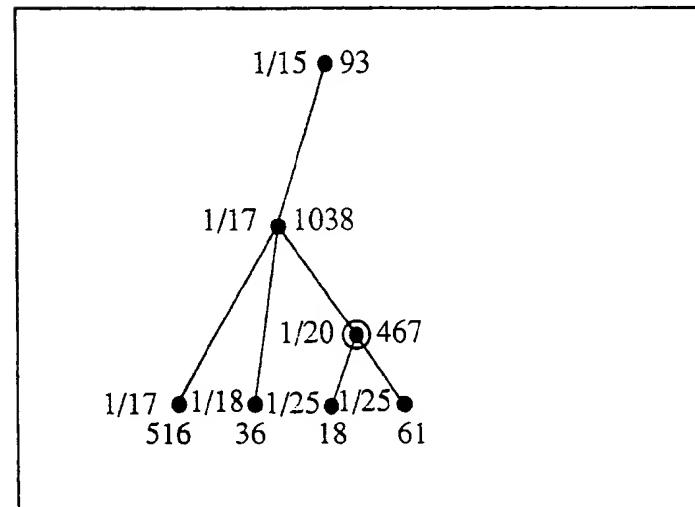


FIG.35B



## FIG.36

DETECTION DATE & TIME		1993 ; 01/25 ; 13 : 58
DETECTION INPUT SENTENCE		A design tool based on examples developed with an object of a computer design.
VIEWPOINT	OBJECT	COMPUTER , DESIGN
	CONCLUSION	EXAMPLE , DESIGN , TOOL
# OF DETECTED DOCUMENTS		61
# OF OK DOCUMENTS		1
OK DOCUMENTS		"Computer design based on examples"
# OF NG DOCUMENTS		2
NG DOCUMENTS		"Control system" "Translation based on examples"

FIG.37

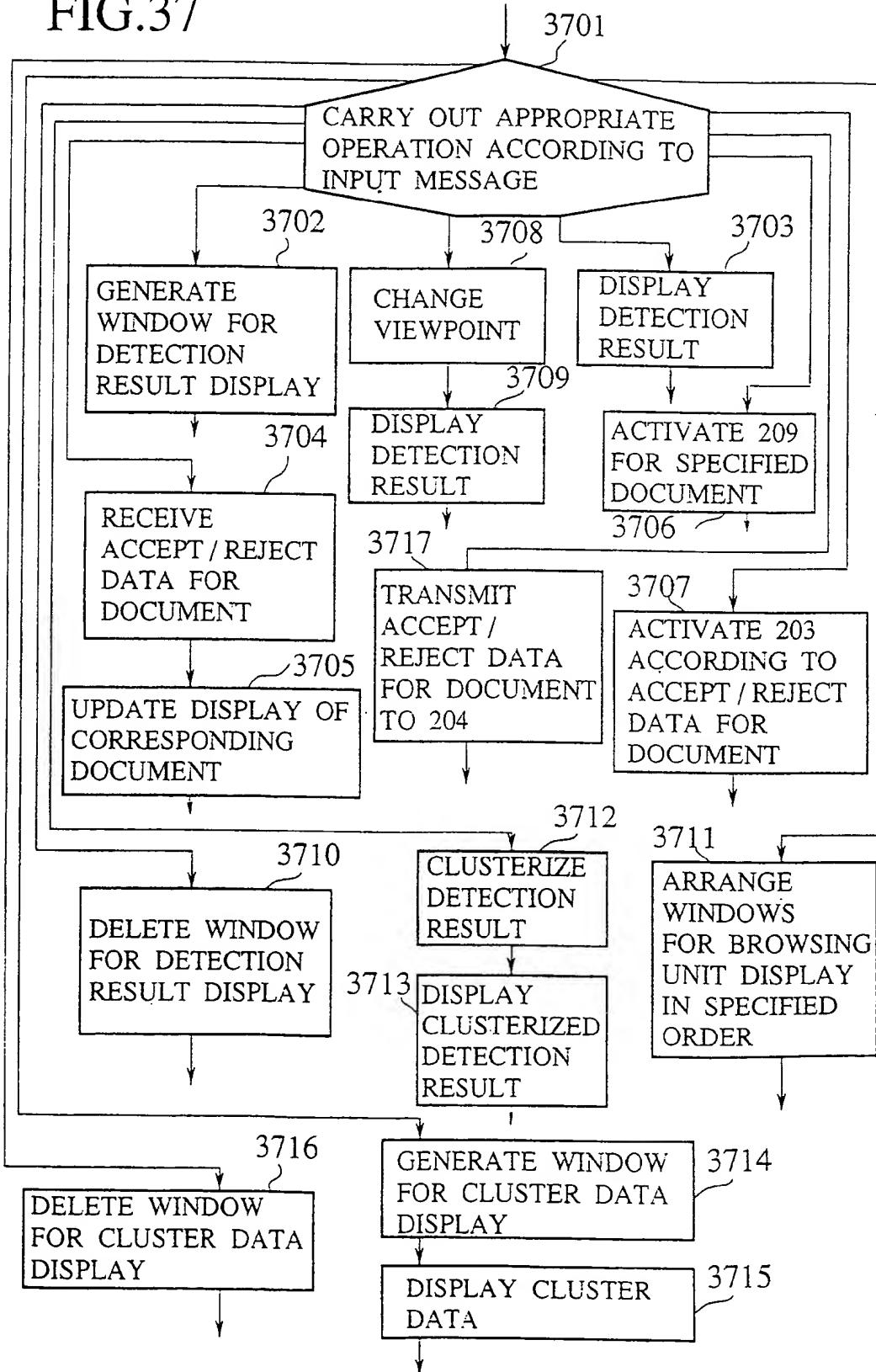


FIG.38

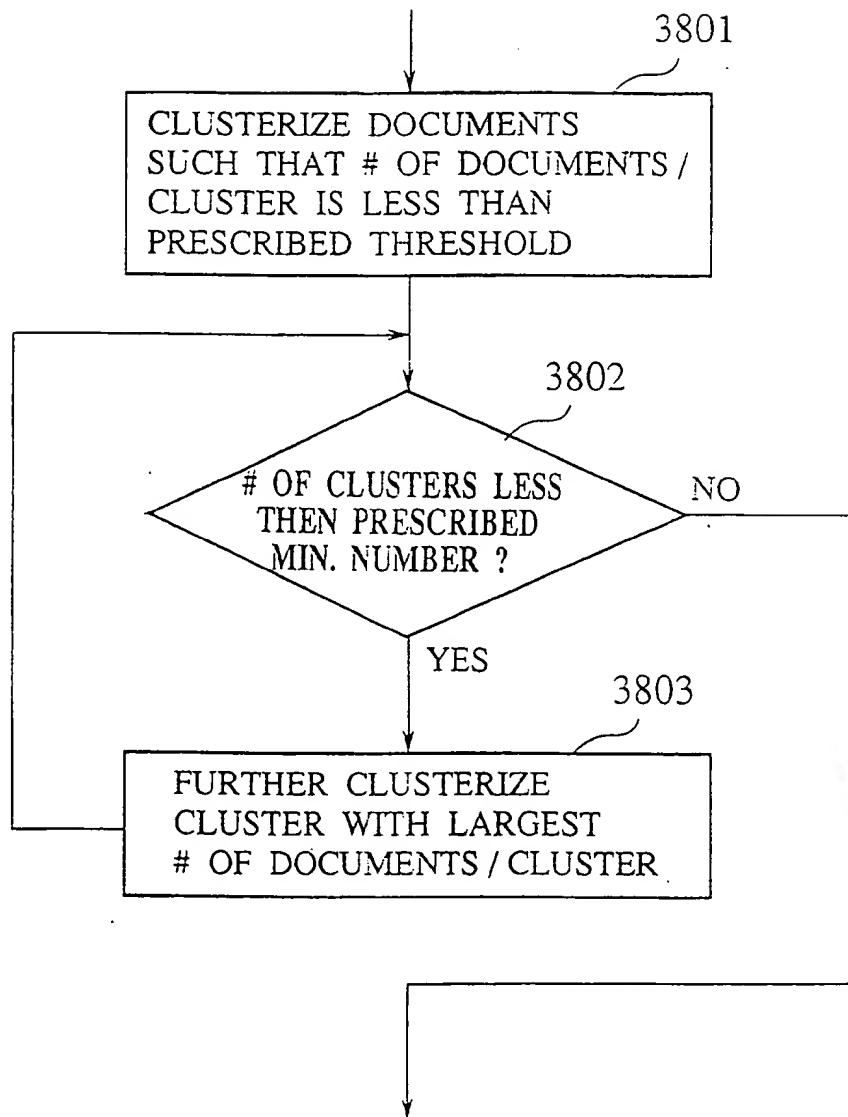


FIG.39

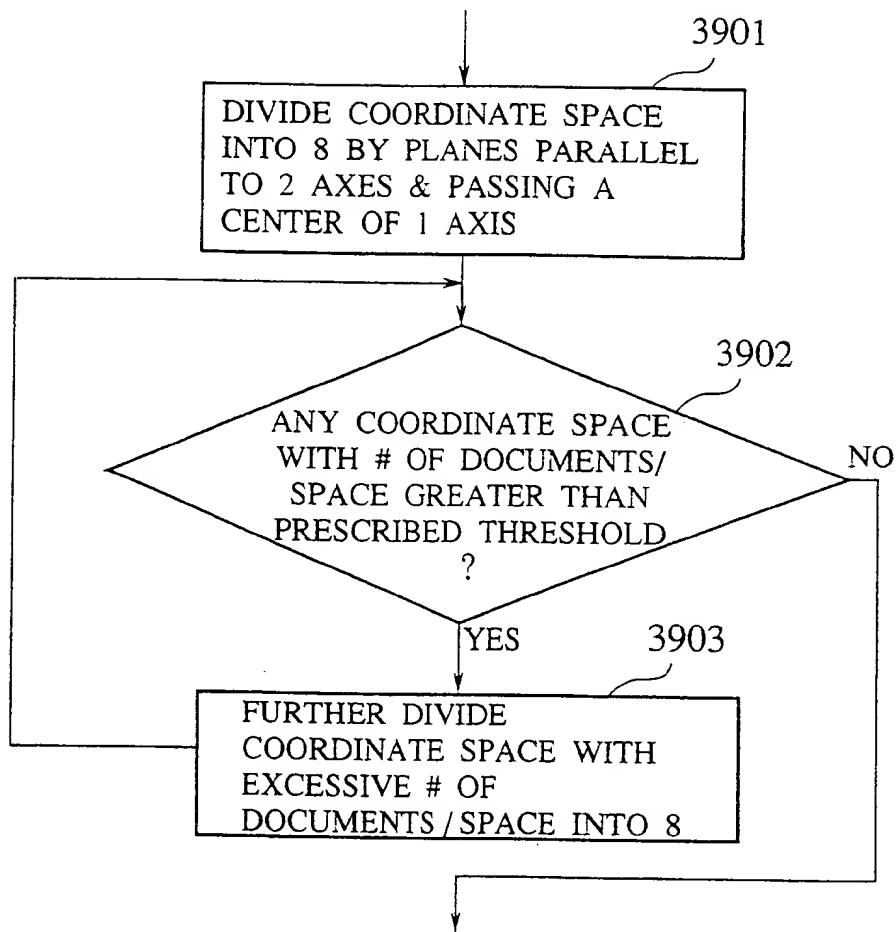


FIG.40

〈FORMAT〉	〈EXAMPLES〉
〈CLUSTER ID〉 ; 〈DOCUMENT ID <sub>1</sub> 〉 , 〈DOCUMENT ID <sub>2</sub> 〉 , . . . 〈DOCUMENT ID <sub>n</sub> 〉 .	Cluster 1 ; file21,file28,file29 file30,file35. Cluster 2 ; file3,file8,file42. :

FIG.41

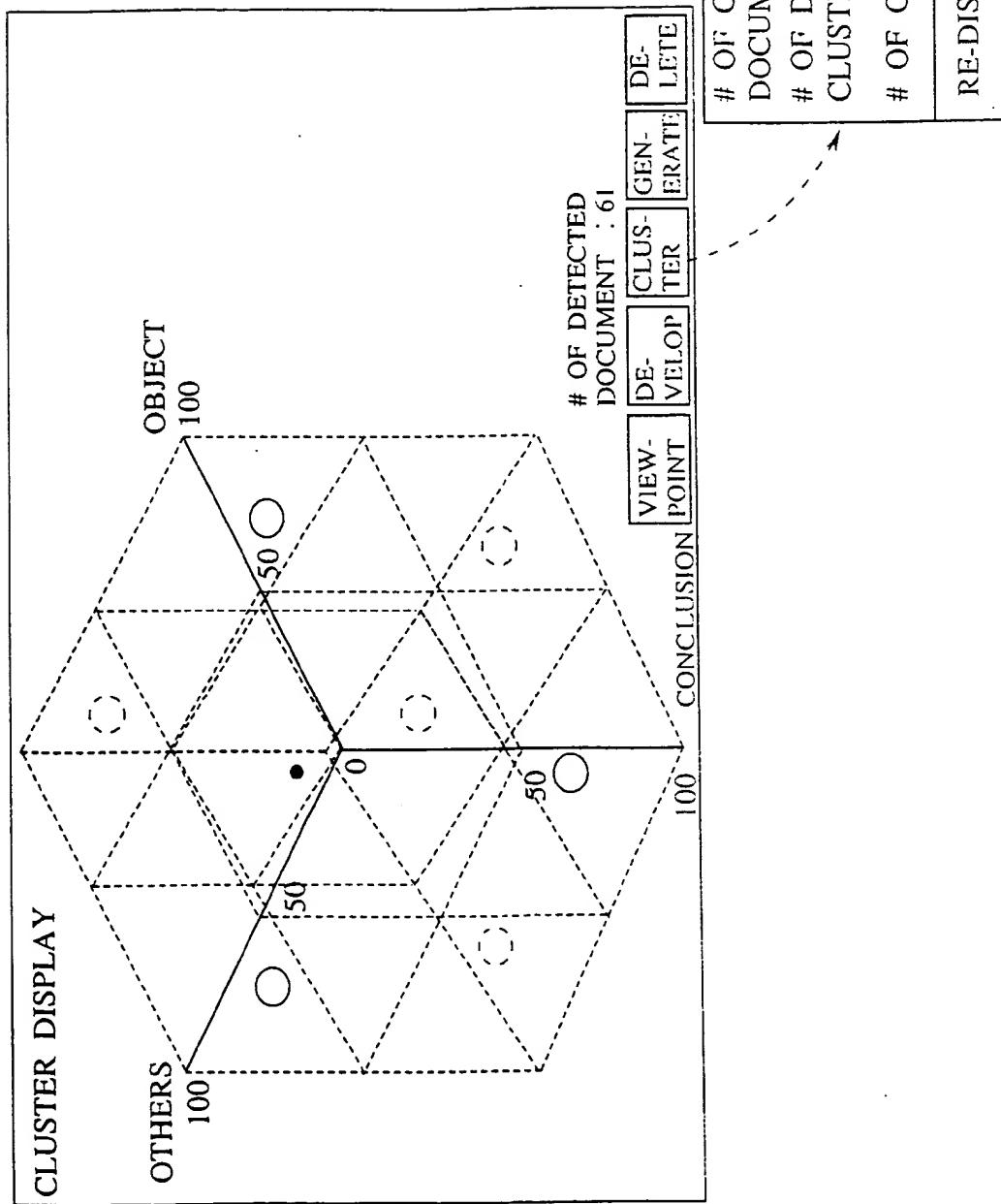
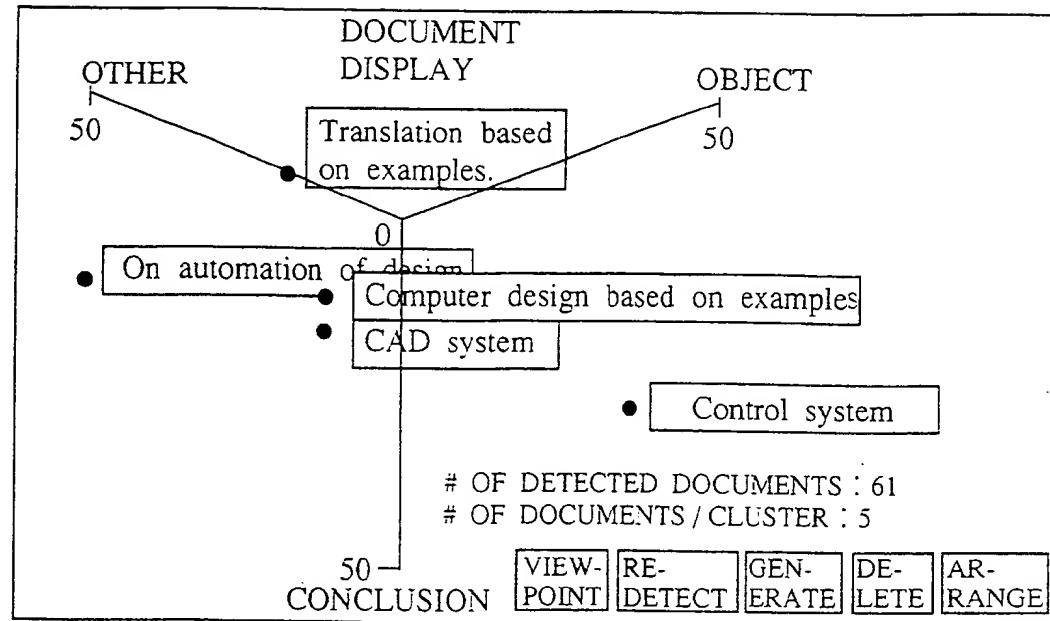
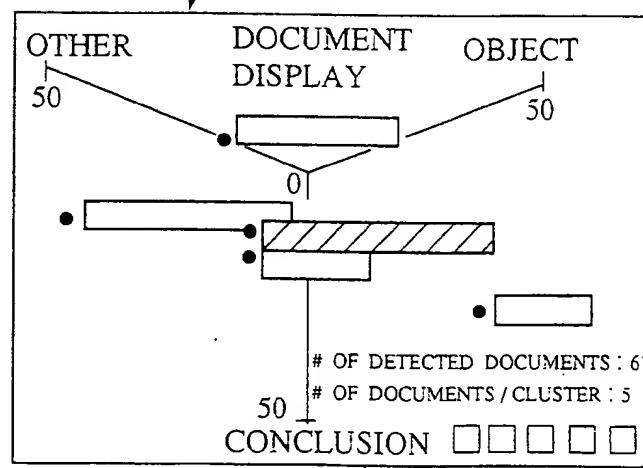


FIG.42

(a)



(b)



(c)

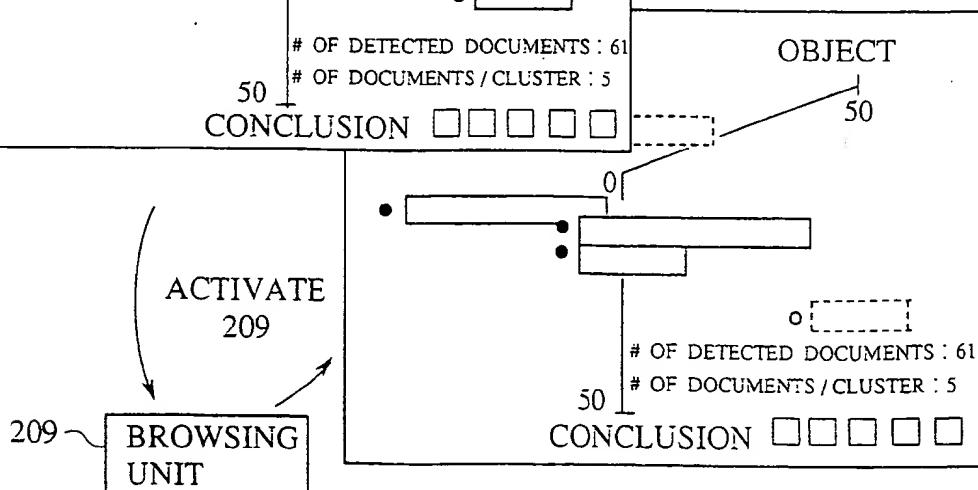


FIG.43

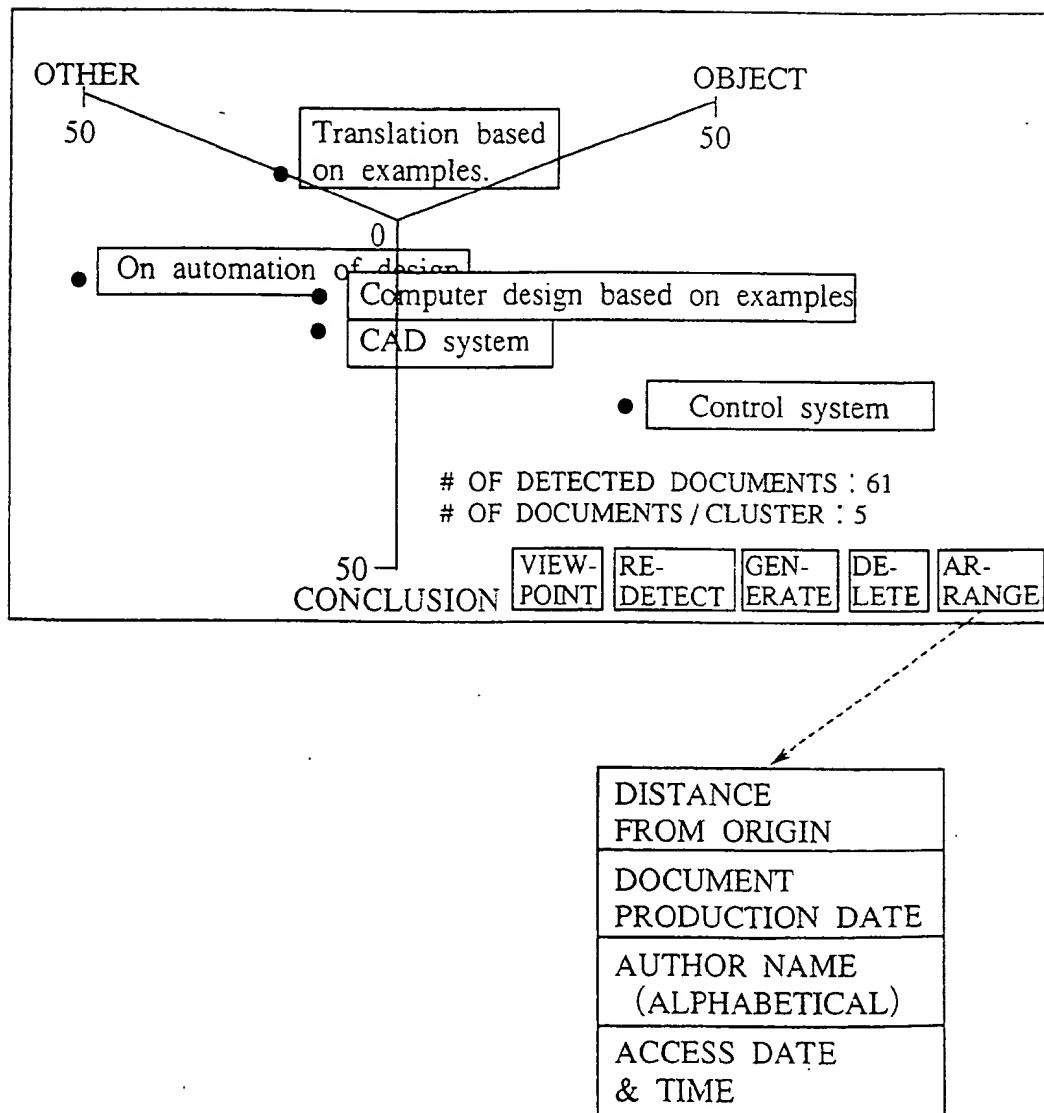


FIG.44A

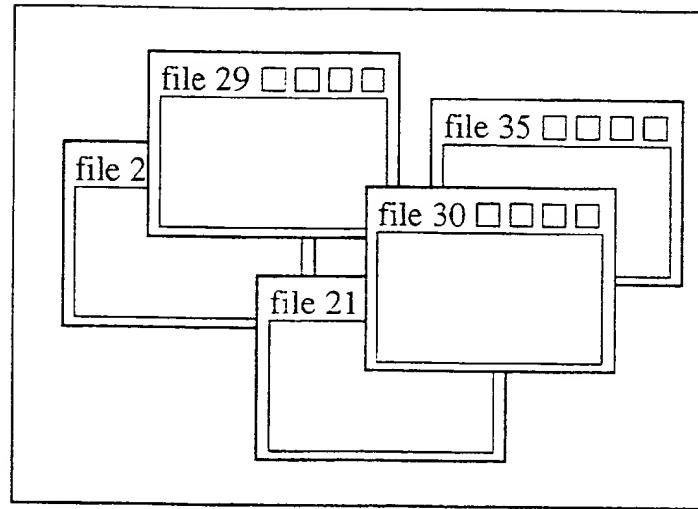


FIG.44B

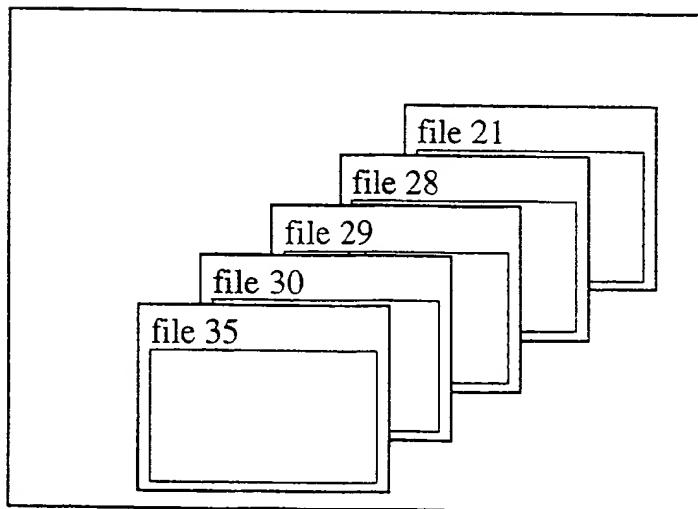
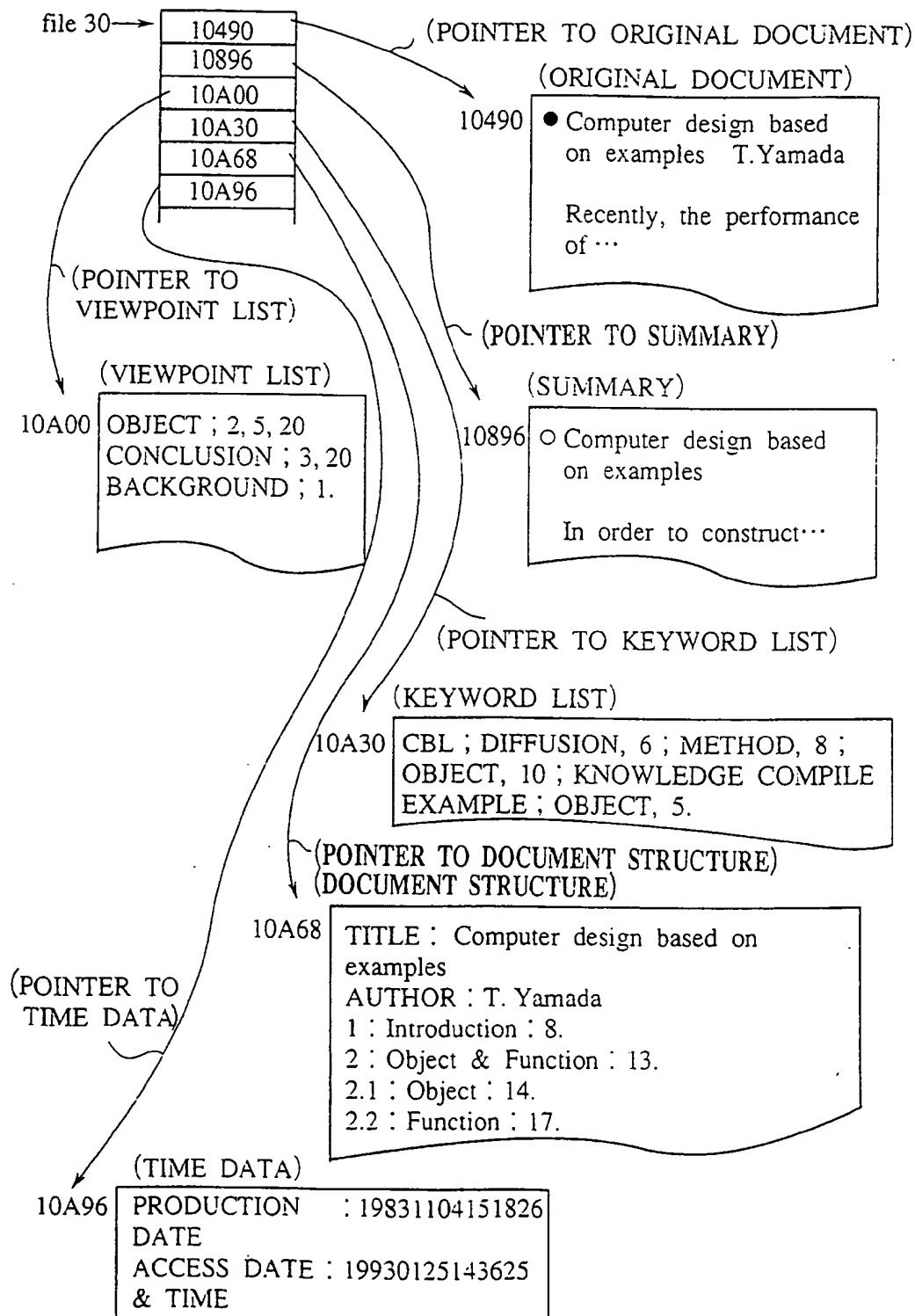


FIG.45



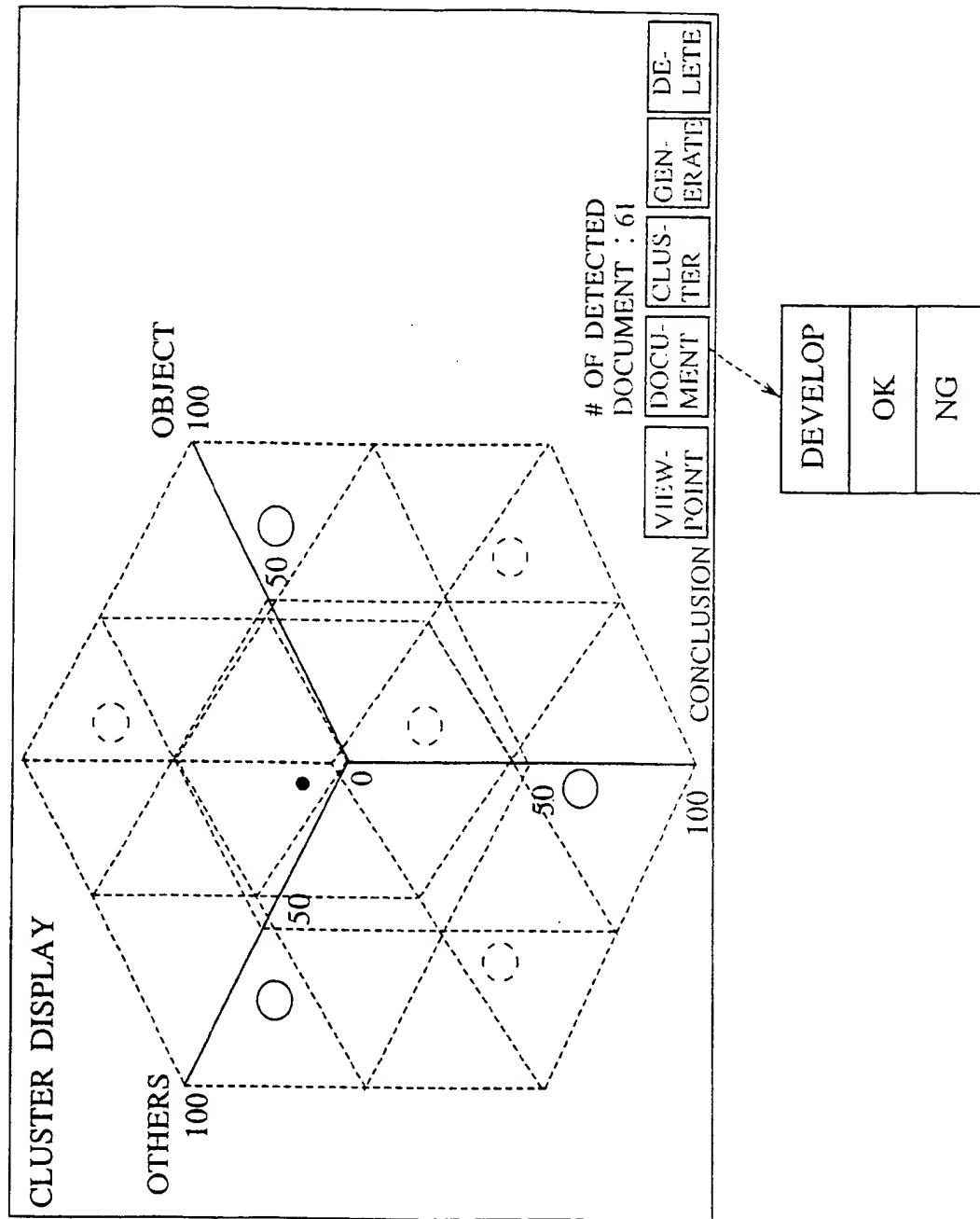


FIG.46

FIG.47

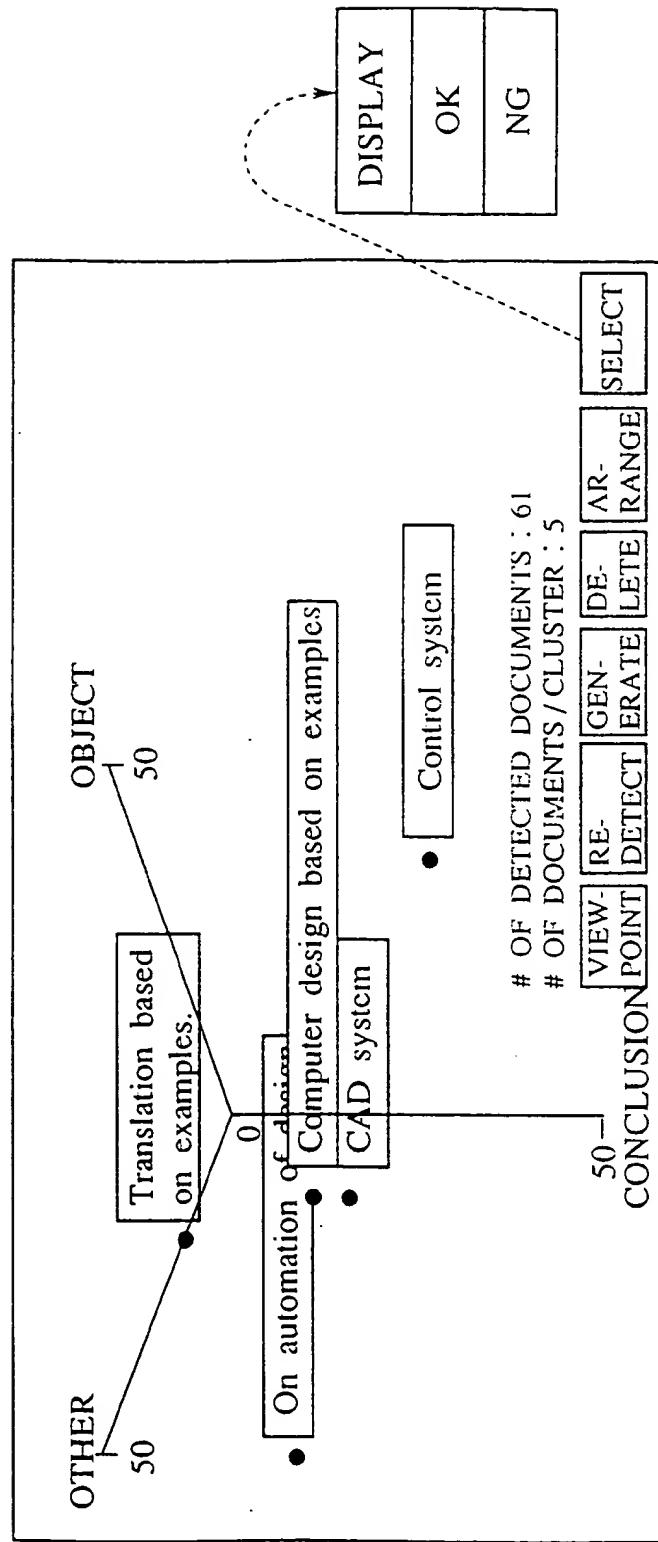


FIG.48

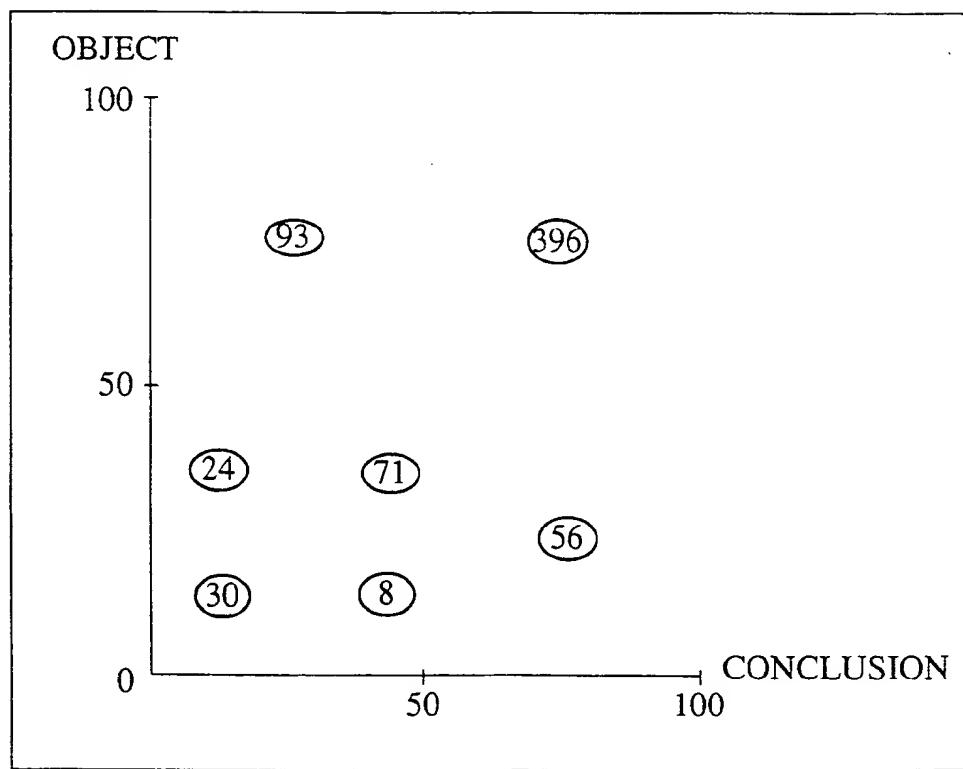


FIG.49

	OB-JECT	CON-CLUSION	BACK-GROUND	METHOD	OTHERS
1	Translation based on examples	0	0	20	60
2	Computer design based on examples	0	0	30	50
3	Research on design automation	0	10	28	50
4	CAD system	0	20	10	30
5	Construction of large scale knowledge base	30	10	67	51
6	Analysis and thoughts on citations	42	29	0	67
7	Data extraction and user adaptation	50	13	78	19
8	Control system	50	15	18	36
9	CASE and knowledge base	60	11	23	51
10					81
11					
12					

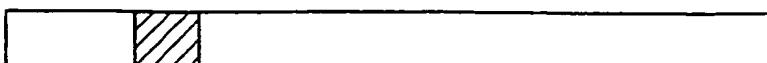


FIG.50

	TOTAL	OB-JECT	CON-CLUSION	BACK-GROUND	METHOD	FEATURE
1	TECHNICAL REPORT DATABASE	3162	3019	3160	2816	3102
2	REFERENCE ABSTRACT DATABASE	1513	1312	1463	721	1096
3	SEARCH REPORT DATABASE	691	689	690	510	613
4	USER MANUAL	57	46	9	6	55
5	DRAWINGS MANAGEMENT DOCUMENT	236	178	12	8	6
6	PRODUCT CATALOG	43	21	13	7	24
						13
						18


  
 1      2      3

FIG.51

(SELL (FROM (MONTH) / (DAY) / (YEAR) ) )  $\Rightarrow$  SALES DATE  
(SELL (FOR PRICE) )  $\Rightarrow$  PRICE  
(IS ((NUMBER) MIPS) )  $\Rightarrow$  PROCESSING SPEED

FIG.52

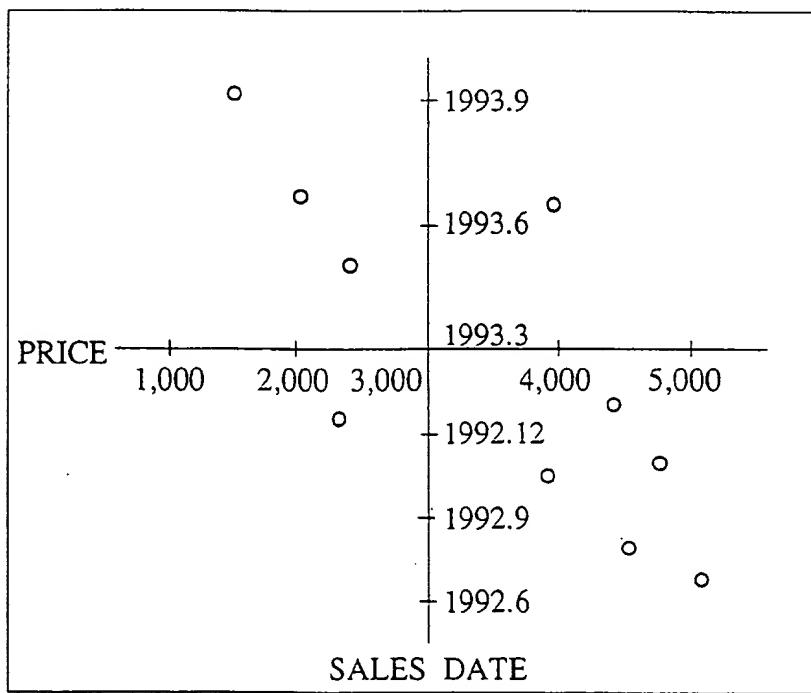


FIG.53A

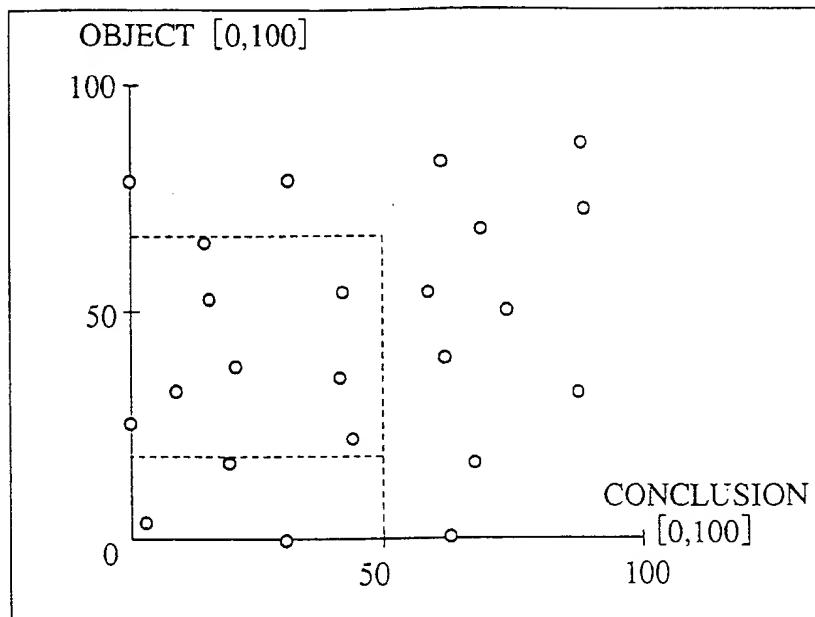


FIG.53B

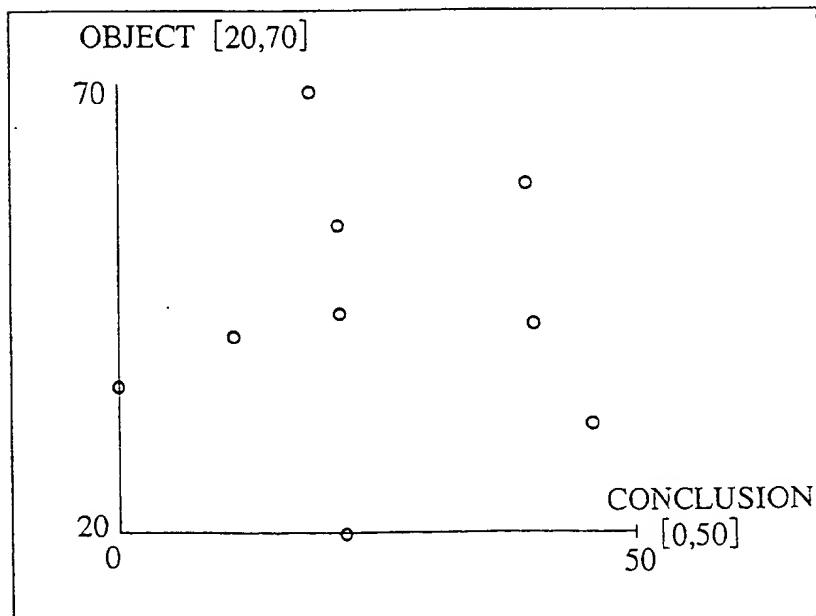


FIG.54

(CONCLUDE [PAST TENSE])  $\Rightarrow$  CONCLUSION, 2

(DEVELOP [PAST TENSE])  $\Rightarrow$  CONCLUSION, 1

FIG.55A

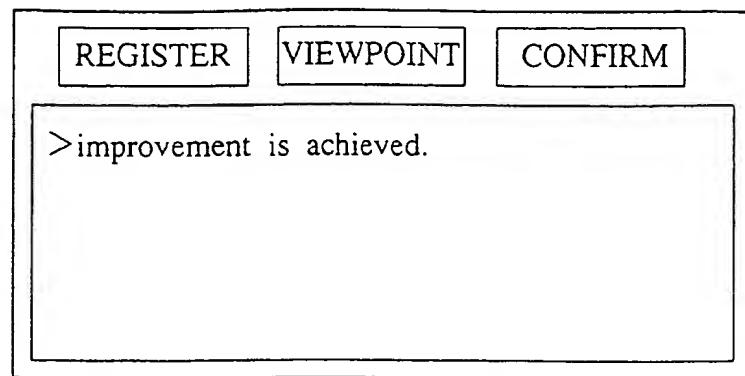


FIG.55B

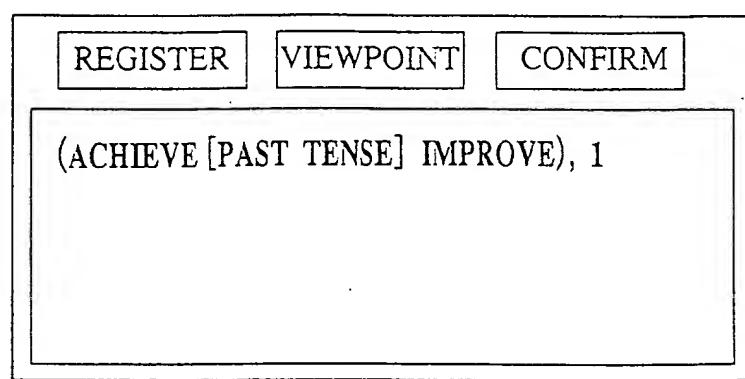


FIG.55C

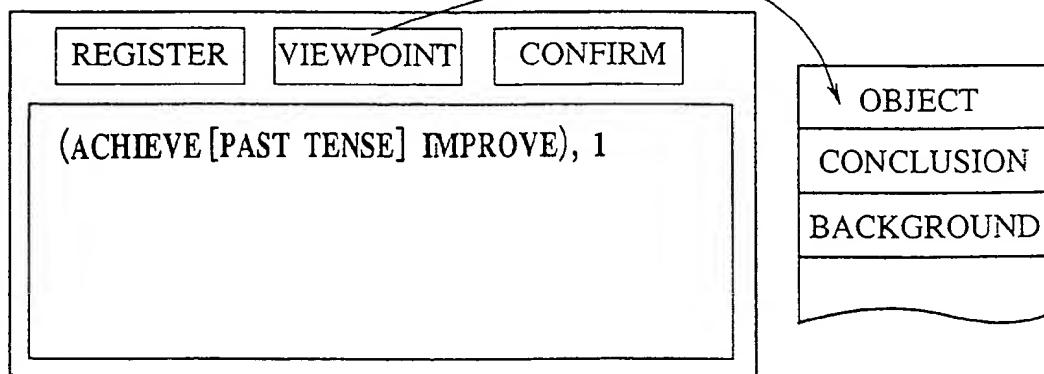


FIG.55D

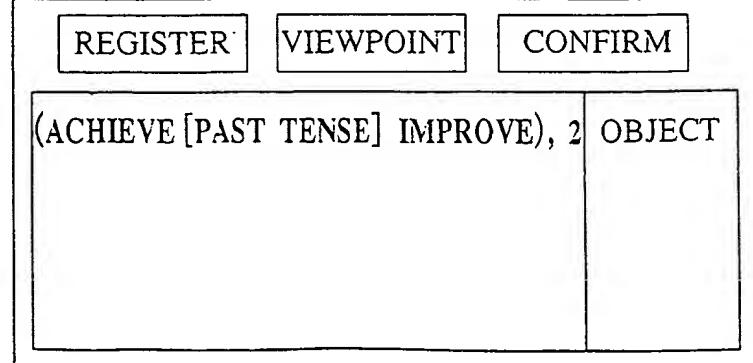


FIG.56A

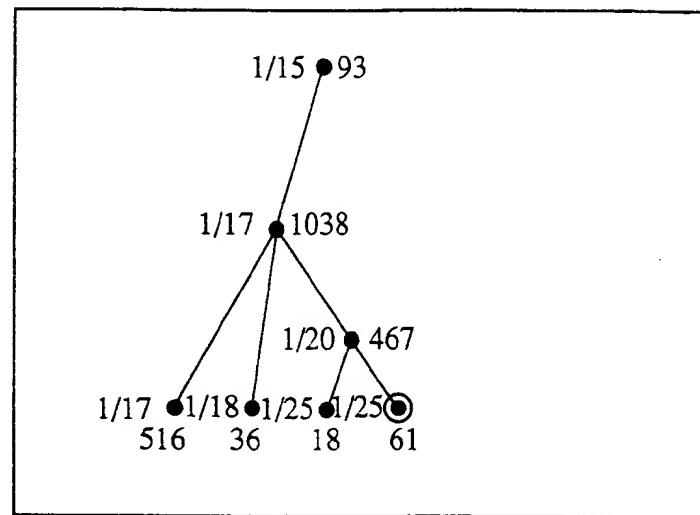


FIG.56B

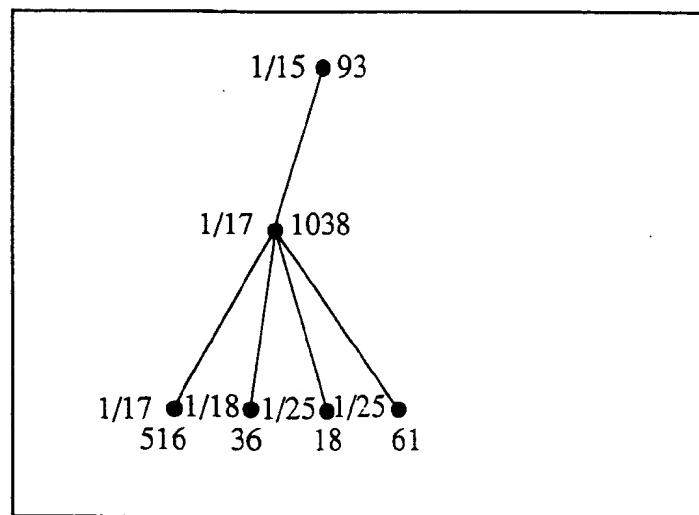


FIG.57

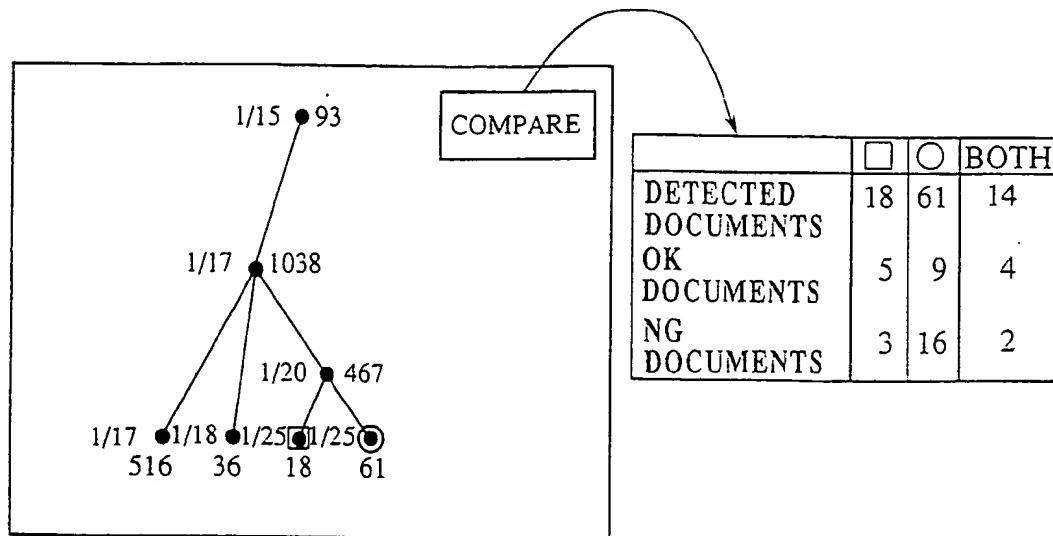


FIG.58

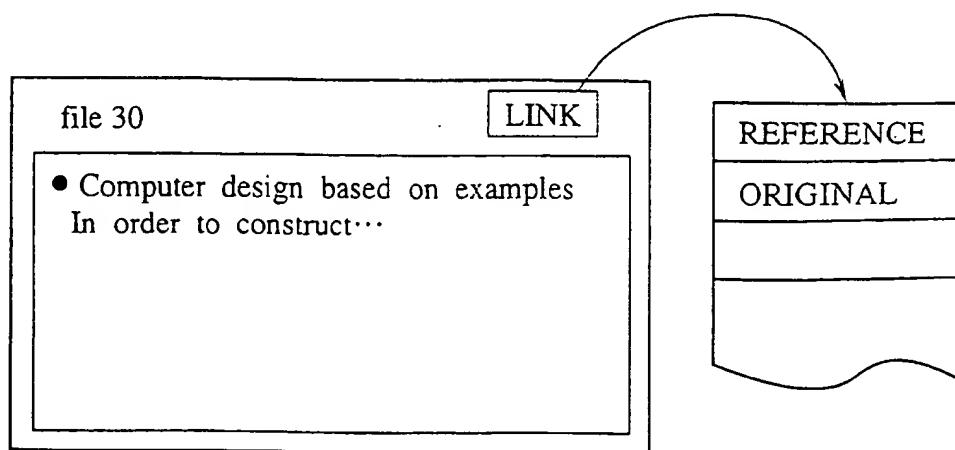


FIG.59

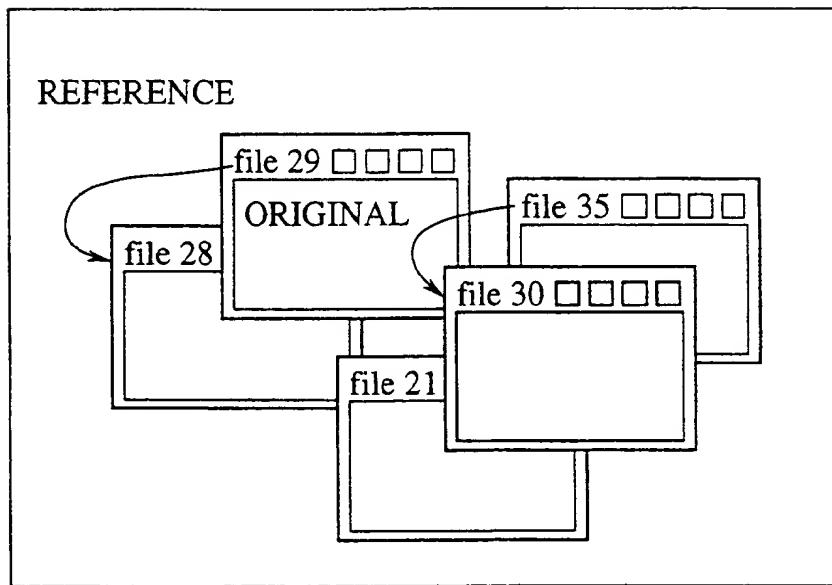


FIG.60

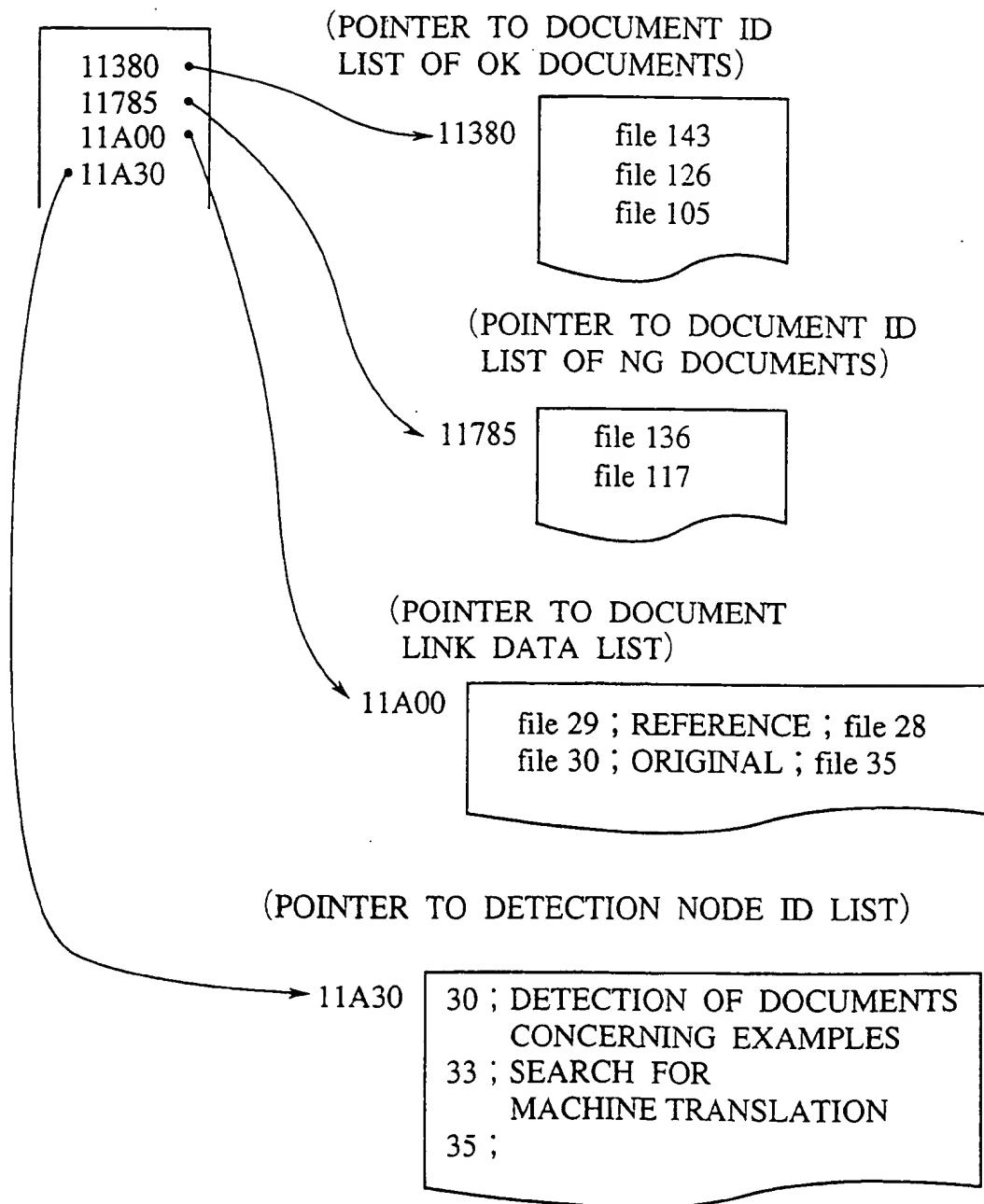


FIG.61

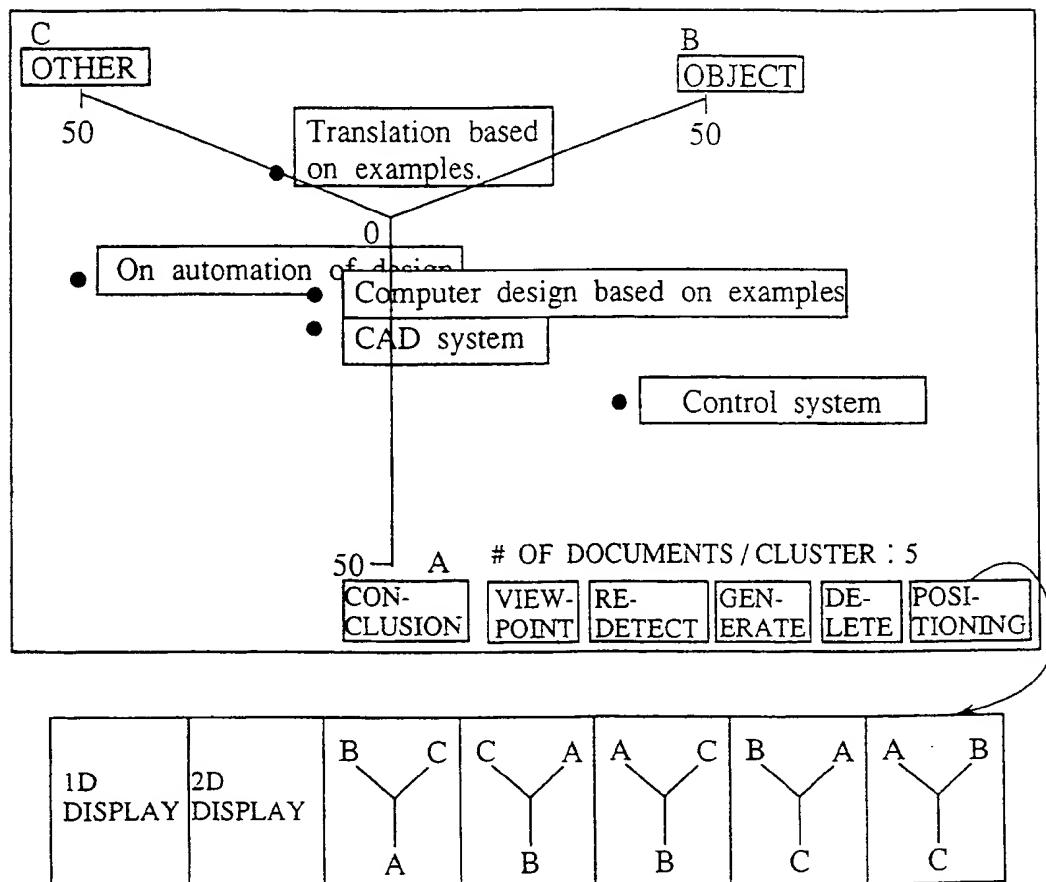


FIG.62

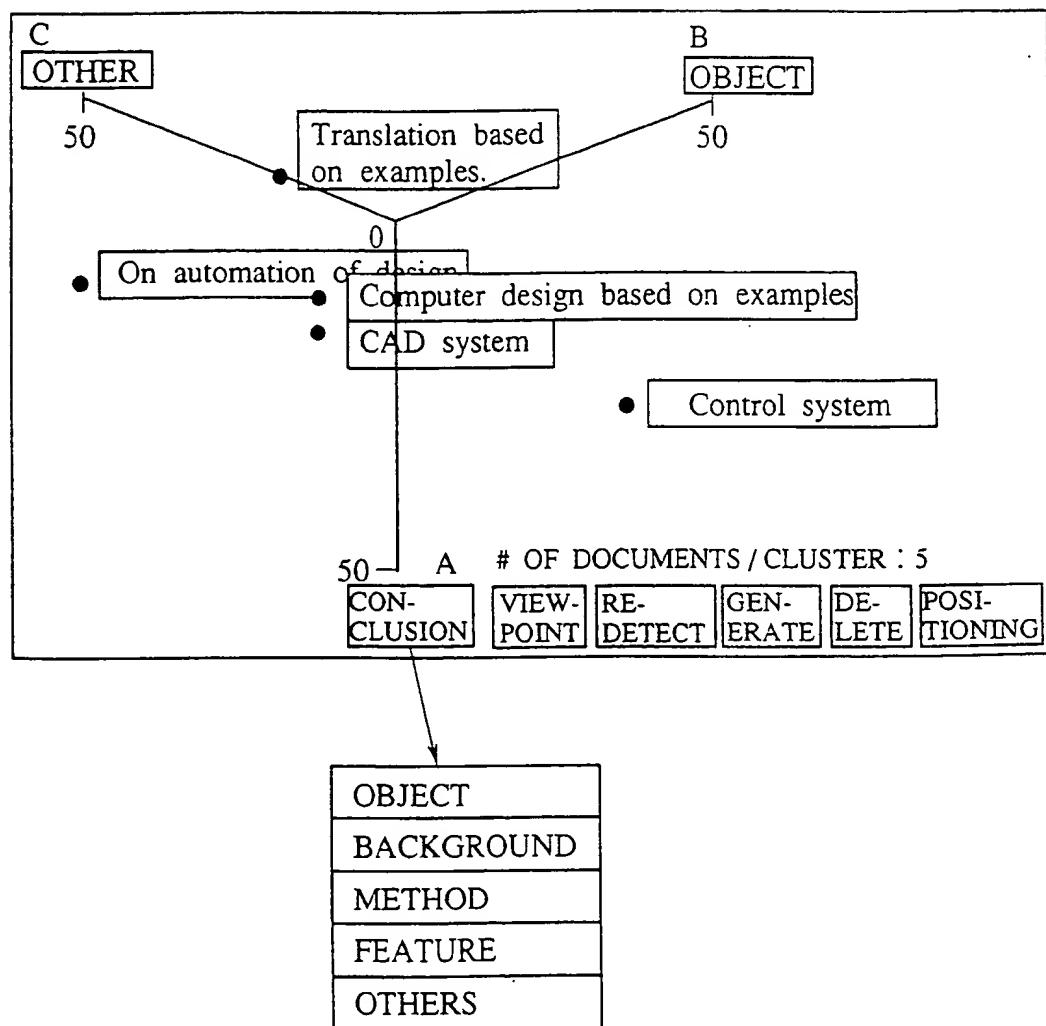
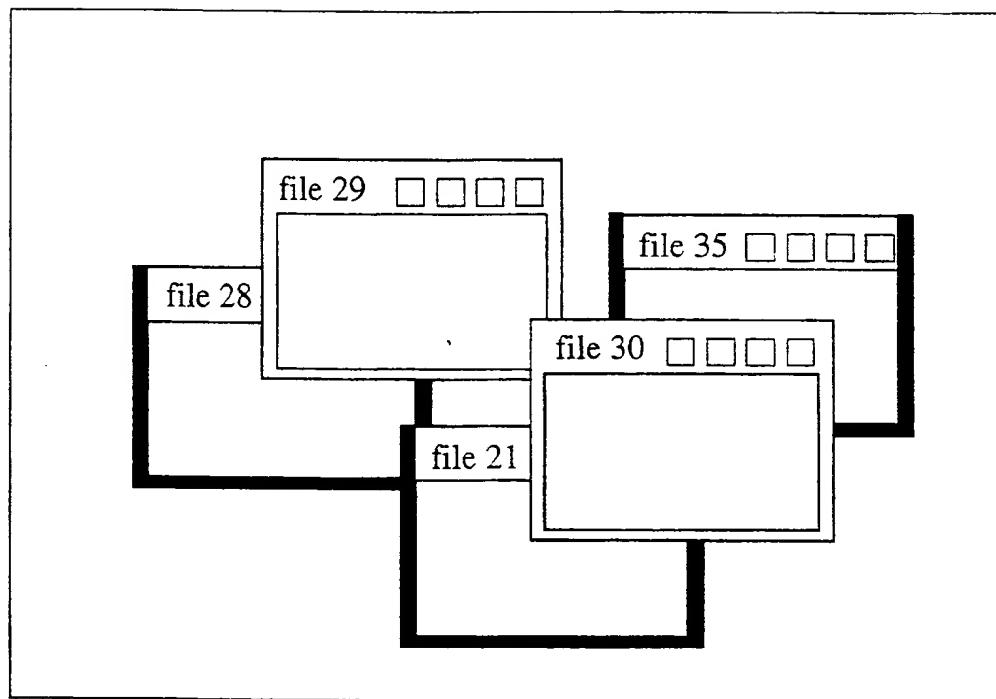


FIG.63





Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) Publication number:

0 615 201 A3

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number: 94103922.4

(51) Int. Cl.<sup>6</sup>: G06F 15/403

(22) Date of filing: 14.03.94

(30) Priority: 12.03.93 JP 79149/93

(43) Date of publication of application:  
14.09.94 Bulletin 94/37

(84) Designated Contracting States:  
DE FR GB

(88) Date of deferred publication of the search report:  
08.02.95 Bulletin 95/06

(71) Applicant: KABUSHIKI KAISHA TOSHIBA  
72, Horikawa-cho  
Saiwai-ku  
Kawasaki-shi  
Kanagawa-ken 210 (JP)

(72) Inventor: Sumita, Kazuo  
67-17, Sachigaoka,  
Asahi-ku  
Yokohama-shi,

Kanagawa-ken (JP)  
Inventor: Hirakawa, Hideki  
1-18-24, Kachida-minami,  
Kouhoku-ku  
Yokohama-shi,  
Kanagawa-ken (JP)  
Inventor: Miike, Seiji  
6-21-19-205, Tsunashima-nishi,  
Kouhoku-ku  
Yokohama-shi,  
Kanagawa-ken (JP)  
Inventor: Mizutani, Yumi  
3-6-7, Yato-cho  
Tanashi-shi,  
Tokyo (JP)

(74) Representative: Lehn, Werner, Dipl.-Ing. et al  
Hoffmann, Eitle & Partner,  
Patentanwälte,  
Arabellastrasse 4  
D-81925 München (DE)

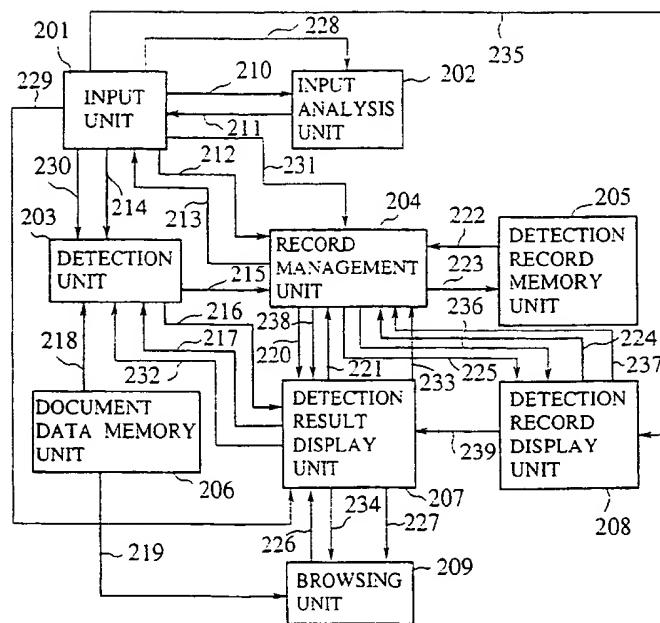
(54) Document detection system using detection result presentation for facilitating user's comprehension.

(57) A document detection system using a detection result presentation that can facilitate a user's quick comprehension of the relevancy of each detected document. In this system, a user's input for commanding a document detection is analyzed to extract keywords and viewpoints relevant to each keyword contained in the user's input, and a detection command is constructed from the keywords and the viewpoints extracted from the user's input. Then, the detection operation is executed to detect those doc-

uments among a plurality of stored documents which match with the constructed detection command as detected documents for a detection result. The detection result can be presented in the multi-dimensional display formed by setting the viewpoints to axes with the detection command as an origin and using distances of the detected documents with respect to the origin for each viewpoint as coordinates of the detected documents with respect to each axis representing each viewpoint.

EP 0 615 201 A3

FIG. 2





European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number

EP 94 10 3922

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)
X	PROC. 14TH. ANNUAL INT. ACM/SIGIR CONF. ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 13 October 1991, CHICAGO, USA pages 270 - 279 WENDLANDT E.B. ET AL. 'Incorporating a Semantic Analysis into a Document Retrieval Strategy' * the whole document *	1,2,18	G06F15/403
Y	US-A-5 020 019 (OGAWA Y. ET AL.) * claim 1 *	4,7,8,19	
A	---	3	
Y	US-A-5 020 019 (OGAWA Y. ET AL.) * claim 1 *	7,8	
Y	PROC. 14TH. ANNUAL INT. ACM/SIGIR CONF. ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 13 October 1991, CHICAGO, USA pages 123 - 132 SUTCLIFFE R.F.E. 'Distributed Representations in a Text Based Information Retrieval System: a New Way of Using the Vector Space Model' * page 125, left column, line 14 - page 129, right column, line 12 *	4,19	
A	RESEARCH DISCLOSURE, no.346, February 1993, HAVANT GB page 156 ANON. 'Extending the Vector Space Model to Include a Reading Level Dimension' * the whole document *	1	TECHNICAL FIELDS SEARCHED (Int.Cl.5) G06F
A	RESEARCH DISCLOSURE, no.346, February 1993, HAVANT GB page 156 ANON. 'Extending the Vector Space Model to Include a Reading Level Dimension' * the whole document *	1,3	
Y	US-A-4 839 853 (DEERWESTER S.C.) * column 2, line 48 - column 4, line 28; figure 1 *	4	
		-/-	
The present search report has been drawn up for all claims			
Place of search  EPO FORM 1502 (02/92) (Postponed)	Date of completion of the search  BERLIN	Examiner  Deane, E	
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons -----         & : member of the same patent family, corresponding document	
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			



European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number  
EP 94 10 3922

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.CLS)
A	PROC. OF THE 14TH. ANNUAL INT. ACM/SIGIR CONF. ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 13 October 1991, CHICAGO, USA pages 134 - 141 KORFHAGE R.R. 'To See, or Not to See - Is That the Query' * page 137, right column, line 28 - page 139, left column, line 4; figure * -----	12-15,20	
			TECHNICAL FIELDS SEARCHED (Int.Cl.5)
The present search report has been drawn up for all claims			
Place of search	Date of completion of the search	Examiner	
BERLIN	23 November 1994	Deane, E	
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons A : member of the same patent family, corresponding document	
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background C : non-written disclosure P : intermediate document			

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**